THE HEALTH AND ECONOMIC EFFECTS OF

DRINKING WATER

By

David McCormick

Cristine Candela

Project Officer

Dennis Tihansky
Washington Environmental Research Center
Washington, D. C.

Prepared for

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

This study examines the relationship between sub-clinical common illness such as upset stomach and other low level diseases, and quality of drinking water and water treatment plant characteristics. Such illnesses account for the greatest occurrence of health problems in the country. For example, in 1969 the National Health Survey found that between 67 and 72 percent of disability days arising from illness fell into the sub-clinical category.

The basic hypothesis tested was that if one accounts for all macro-systemic influences on health, the occurrence of sub-clinical illnesses will be highest in those communities with the worst water quality. The measures of drinking water quality and water treatment plant character-istics were obtained from a Public Health Survey study entitled, "Commu-nity Water Supply Study". Elementary school absenteeism was used as a surrogate measure for sub-clinical disease. The primary statistical tool employed was stepwise multiple regression.

A linear regression was performed of elementary school absenteeism onto total plate counts, measures of water treatment facilities, and a series of control variables. The resulting regression was essentially a damage function from which percentage absences resulting from incre-mental changes in water quality were predicted. Under the assumption that low level illness in the adult population is affected by a constant proportion to childrens' illness, the number of drinking water-related absences and their associated dollar value were estimated.

iii

CONTENTS

SECTION I

CONCLUSIONS

A consistent relationship was found between operator salary and elementary school absenteeism, controlling for community wealth and urbanization. The frequency of testing finished water, times the salary of the operator(s), results in a useful index of the operation of a water treatment facility.

Regression analyses allowed us to predict the percentage of elementary school absences which result from incremental changes in water quality. The results broadly indicated that a .2 to .5 percent increase in school attendance can be expected for each $1000 increase in operator salary. Nationwide and extrapolating to the adult population, improvements of this magnitude would produce savings (in wages and medical expenses) of approximately $250 million per year.

Conclusively, our findings show that water treatment plant operators' salaries appear to be a useful predictor of elementary school absenteeism. However, one cannot assume that pay raises will immediately improve community health. The very low salaries which are now prevalent have resulted in the recruitment of poorly educated and undertrained operators. While, in time, higher salaries will make possible the selection of higher caliber operators, in the short term, the salaries would have to be supplemented by vigorous operator improvement programs if any significant result is to be expected.

Therefore, we computed benefits on the basis of increments in improvements in the various measures of water quality. The various increments predict reductions of 3% to 10% of the digestive and non-respiratory illnesses. In the specific areas studied, this translated to reductions in elementary school absences on the order of 300,000 to 600,000 absences per year per incremental improvements. Converting this to worker absenteeism using a constant of proportionality of .62, this becomes a savings in the order of 23 million dollars per year per increment. For the nationwide sample, we would cautiously estimate that elementary school absences would be reduced on the order of magnitude of 11-12 million absences per year per incremental im-provement. The wages and medical expenses saved would range in the order of 250 million dollars per year per incremental improvement.

SECTION II

RECOMMENDATIONS

In carrying out the study, every attempt was made to be conservative in our estimates. The medical cost figures, for example, assume that there are no respiratory diseases, resulting from drinking water. If they were transmitted in a manner similar to gastrointestinal illness, the medical cost figures would be more than twice what we have estimated. At all other points in the research program, when we were faced with two alternatives, one of which would tend to inflate the results and the other of which would tend to deflate them, we systematically chose the more conservative.

We believe that the results indicate that if improvements are made in the quality of drinking water, substantial savings will be realized. As the CWSS report suggested and our study has confirmed, the problems are most severe in the rural non-farm communities. There the economic savings alone would be most dramatic both because the effects are relatively large, and the costs of resolving the problem are relatively small. In most cases, upgrading the personnel and procedures appears to be all that is needed.

SECTION III

PROBLEM AND APPROACH

Background

Conventional rehtoric used to be that U.S. drinking water was of the highest quality and purity. In fact, most Americans took water purity for granted, considering it a fact of life. Recent studies, however, have served to dispel such inaccuracies. Besides the alarming discoveries of carcinogenic agents in drinking water, there have been analyses such as Craun and McCabe's findings that 1) water-related epidemics affecting an average of 100 persons per outbreak occur at an average rate of one per month;[1] 2) McDermott's estimates that of annual occurrences of 500,000 infections of hepatitis and two million of aseptic meningitis, both can be transmitted by water[2]; and 3) Mosley's 1967 survey of waterborne disease viruses, which opened a wide range of possible links to major health problems.[3]

Community Water Supply Study, 1970

In July 1970, the Department of Health, Education, and Welfare published results of their own nation-wide evaluation of drinking water quality and water treatment plants. This Community Water Supply Study (CWSS) shows that the quality of drinking water is far below general expectations.[4] Significant results were:

- Forty-one percent of the water systems, delivered inferior water to 14% of the study population.

- More than one-third of the tap water samples contained contaminants in excess of Public Health Service standards.

- Over half of the systems had physical or operational deficiencies.

---

1) G. Craun, L. McCabe, "Waterborne Disease Outbreaks, 1961-1970". Annual Meeting of the American Water Works Association, 1971.

2) J. H. McDermott, "Virus and Water Quality: Occurrence and Control Conference Summary" (13th Water Quality Conference, February 1971), p. 213.

3) Mosley, James, "Transmission of Viral Diseases by Drinking Water" from Bery (eds.) Transmission of Viral Diseases by the Water Route, New York, 1965.

4) McCabe, Leland J., J. Symons, R. Lee, and G. G. Robeck, "'Survey of Community Water Supply System", Journal of the American Waterworks Association, Vol. 62, No. 11, November 1970, p. 670-687.

- More than three-fourths of the plant operators had inadequate training in fundamental water microbiology; nearly half in application chemistry.

- Control, inspection, surveillance, and testing programs were entirely lacking, infrequently performed, or well below standards in the vast majority of systems.

Surprising as they may be, the CWSS results are probably optimistic, since water quality measurements and standards suffer from a lack of virus data. Direct methods for measuring viruses in large quantities of water are under development but not yet available. Although present water purification technology can handle all significant bacterial problems, in practice, available technology is not always applied. And it is far from certain that viruses and harmful chemicals can be easily removed with existing technology.

In recent years, the American press and popular literature have reflected a growing public awareness of environmental pollution. At the same time, public attention has focused on the cost in higher taxes and consumer prices to clean it up. Confronted with this price tag, taxpayers and the decision-makers who represent them naturally ask what they're buying.

In the area of water pollution control, there are numerous payoffs, including direct economic benefits such as commercial fishing, measured real estate values, increased recreational use, and generally improved aesthetic value, all of which contribute to improving quality of life for Americans.

Among the costs incurred by our degraded water supplies are those of purifying it to make it safe, to drink. Alternatively, of course, failure to purify the water sufficiently can be argued to increase the possibility of illness.

However, the question arises whether and to what extent incremental degradation of the drinking water has negative effects on community health. If such effects exist, there are associated, definitive economic implications both on the cost of operating the treatment facilities and on the costs of improving the quality of the intake water quality.

This Study

Numerous studies have already been undertaken to examine the effects. of water degradation on serious communicable disease. In view of this fact, this study proposed an examination of data relating to sub-clinical illnesses such as gastroenteritis and other non-specific low level diseases. Such illnesses account for high occurrence of health problems in the United States. In 1969, for example, the National Health Survey found that between 67 and 72 percent of disability days arising from illness fell into the sub-clinical illness category.

Based on the assumption that the economic costs of sub-clinical illnesses affect a far wider distribution of people than do major illnesses, lost productivity alone due to sub-clinical absences in 1969 totaled in the

4

general range of $4 1/2 billion dollars.* Consequently, the costs of sub-clinical illness are large and, in fact, are more than twice the total EPA 1974 estimated budget outlays.

This study was, therefore, undertaken with the basic hypothesis that any significant health effects from water degradation would result primarily in sub-clinical illnesses. To perform the study, some reliable measure was needed for community health. Worker sick time data, the most direct measure are not usually reported for a community. Even where they are available, sick time data are notoriously unreliable and can tend to follow a form of Parkinson's Law: sick days off expand to absorb sick days allowable. Therefore, a widely reported indicator of community health had to be chosen. We selected absenteeism from elementary schools as our indicator for the following reasons:

- Records in most school districts are reasonably accurate and complete.

- Elementary school children represent a wide spectrum of socio-economic classes in a community.

- Truancy is less serious problem for early grades, than for high school.

- Both children and adults are susceptible to low level illnesses; therefore, school absenteeism can be linked to work absenteeism.

- Biases in absenteeism due to non-water related phenomena can be partly explained by measurable variations in socio-economic community characteristics.

For the communities studied, we performed a linear regression of elementary school absenteeism onto total plate counts, measures of the treatment facilities, and a series of control variables.

The resulting regression is essentially a damage function from which we were able to predict the percentage absences resulting from incremental changes in water quality. Under the assumption that low level illness in the adult population is affected by a constant proportion to childrens' illness, the number of drinking water related absences and their associated dollar value were estimated. This was done for the nine areas used in the CWSS study: the State of Vermont, and eight standard metropolitan statistical areas (New York, N. Y.; Charleston, W. Va.; Charleston, S. C.; Cincinnatti, Ohio; Kansas City, Missouri-Kansas; New Orleans, Louisiana; Pueblo, Colorado; and San Bernadino-Riverside-Ontario, California. In

---

* Based on 368 million disability days to persons 14 and over due to acute illnesses, 67% of which are sub-clinical, a median, daily wage of $21.00, and 58% employment.

addition, extremely general estimates were made for the total nation.

The ensuing sections of this report will review:

1.) the findings from a literature review covering research on water quality relationships to low level illness;

2.) the research design and methodology, and

3.) description of analysis and findings.

Five appendices list data resources. A bibliography provides references relevant to the research.

SECTION IV

LITERATURE REVIEW

Human Health Effects from Water Pollution

Method - To ascertain what is already known about the costs of water pollution with relation to low level illness, a survey of the relevant literature was conducted. Although several articles are discussed and quoted with regard to water quality and illness, the major studies reported were reviewed critically for the utility of their objectives, fulfillment of these objectives, and the analytical methods used. Sources are mentioned wherever given. Emphasis is given to consistent findings between water quality and illness, with attention to geographical aspects involved.

Related Research

There is still a paucity of research in the area of water and human health, particularly with respect to low level illnesses. Also, there is a dearth of records and statistics dealing with minor illnesses. In this literature review, the goal is to complement our research objective to measure the economic costs to society in lost wages and medical expenses arising from low level illness associated with water quality. The inadequacy of background data is significant. Krishnaswami has pointed out, "that man's direct and indirect contact with water and the aquatic environment varies considerably, and that the total impact of all environmental factors - air, water and food - cannot be compartmentalized in assessing their significance to health and disease". [5] In a quote from Dubos, it is stated that: "Experimental and epidemiological studies (e.g., common cold and infectious hepatitis, respectively), have shown that even in the transmission of pathogenic organisms and the production of overt disease . . . a simple relationship of causal specificity does not always exist. The presence of a pathogen or a toxicant is only a necessary condition, but not a sufficient condition for the causation of a disease." [6]

---

5)   S. Krishnaswami, "Health Aspects of Water Quality" American Journal of Public Health, 61:2260 (November 1971). For a literature search of published and unpublished estimates of the cost of air pollution damage, see Larry B. Barrett and Thomas C. Waddell, Cost of Air Pollution Damage: A Status Report, EPA, February 1973. Also see literature review contained in Lester B. Lave and Eugene P. Seskin "Air Pollution and Human Health", Science 169, 21 August 1970, 723-733.

6)   Dubos, Rene'. Man, Medicine and Environment. New York: The New American Library, Mentor MY 942, 160 pp. 1969.

Krishnaswami adds that, "The overt clinical manifestation of severe chemical deficiency or of acute toxicity can be evaluated and their causes discussed relatively easily. However, the role of water constituents (relatively significant concentrations of various synthetic organics, heavy metals, and other chemical toxicants) in the probable multifactorial causation or contribution to chronic ailments and other bodily stresses have not received serious consideration."[7]

Thus, in conducting studies in the area of water quality and minor illness one must deal with the fact that their occurrence is negligibly reported. Many individuals will forego any physician consultation during low level illnesses, but additionally, records that do exist are incomplete as to causation, circumstances, length of illness, etc.

Waterborne Disease Outbreaks

Weibel, et al.[8]

Despite such limitations, Weibel, et al, were able to compile diseases by identity finding that the number of reported epidemics and cases of waterborne disease outbreaks in the United States between 1946 and 1960 were primarily classified as "gastroenteritis". (See Table 1.) "This category and 'diarrhea' (16 outbreaks and 5,160 cases) are not specific diseases, because the etiologic agent has not been determined through laboratory analysis. Even with laboratory analysis, the causative agent cannot always be determined."[9]

---

7)   S. Krishnaswami, p. 2261.

8)   S. R. Weibel, F. R. Dixon, R. B. Weidner, and L. J. McCabe, "Waterborne Disease Outbreaks, 1946-60", Journal of the American Waterworks Association, 56:-949 (August 1964).

9)   S. Krishnaswami, p. 948.

TABLE 1[10)]

Waterborne-Disease Outbreaks in the United States,

1946-60, by Type of Illness and System

Illness	Private or Semi-public Systems		Public Utilities		All Systems	
	Outbreaks	Cases	Outbreaks	Cases	Outbreaks	Cases
Gastroenteritis	92	4,233	34[11)]	9,397	126	13,630
Typhoid	33	403	6	103	39	506
Infectious hepatitis	14	430	9	500	23	930
Diarrhea	7	320	9	4,840	16	5,160
Shigellosis	4	596	7	5,057	11	5,653
Salmonellosis	3	22	1	2	4	24
Amebiasis	2	36	0	0	2	36
Other	3	16	4	29	7	45
Total	158	6,056	70	19,928	228	25,984

10)   Weibel, et al.

11)   One gastroenteritis outbreak also included a typhoid case.

Weibel, _et al_ also report on a review of waterborne-disease outbreaks in the United States and Canada by Gorman and Wolman,[12] covering the period from 1920 to 1936, in which the case distribution (116, 000) by type of illness was: 88 percent diarrhea, 11 percent typhoid fever, and 1 percent dysentery. A second review by Eliassen and Cummings[13] covered waterborne outbreaks in the United States between 1938 and 1945 with a case illness distribution (110,000) of: 91 percent gastroenteritis (which probably corresponds to the "diarrhea" classification of the earlier study); 8 percent dysentery; and, 1 percent typhoid.

The Eliassen and Cummings report used the term "outbreak" to refer to two or more cases; the Gorman and Wolman report defined it as consisting of five or more cases.

Weibel, _et al_ also show a distribution of outbreaks by month for the years 1946-60, which indicate a peaking of outbreaks due to private supplies in the summer months. (See Table 2.)[14] This is of extrinsic interest to our research in that our stress is rather on usage of school absenteeism data, covering the time periods of September to June. It is a finding which is given corroboration by more recent analyses done by Petersen and Hines, which will be reported on in a later section of this report.

Unfortunately, the only explanation offered with regard to causal factors of such occurrence is either greater pollution during the summer months, or possibly increased usage by more susceptible individuals (tourists, campers, travelers, etc.).

G. Craun, L. McCabe[15]

During 1961-1970, Craun and McCabe reviewed 128 known outbreaks of disease or poisoning attributed to drinking water, with 46, 369 illnesses and 20 deaths. Their definition of "outbreak" reflected at least two cases of infectious disease, associated with water used for drinking or domestic

---

12) Gorman, A. C. & Wolman, A., "Waterborne Outbreaks in the United States and Their Significance. Journal of the American Water Works Association 31:222 (February 1939).

13) Eliassen, R. & Cummings, R. H. "Analysis of Waterborne Outbreaks, 1938-45" Journal AWWA, 40:509 (May 1948).

14) Weibel, _et al_, p. 954.

15) G. Craun and L. McCabe, _Ibid_.

TABLE 2 [16)]

Seasonal Distribution of Waterborne-Disease

Outbreaks in the United States, 1946-60, by

Type of System and Affected Population

Month	Public Utilities %	Private or Semipublic Systems				
		Percent-age of Total	Affected Population			
			1*	2†	3‡	4§
January	7. 6	3. 9	3		2	
February.	10. 6	3 .9	4		1	1
March	7. 6	2 .6	3		1	
April	7. 6	5 .9	4		4	
May	6. 1	6. 6	3		5	
June	3. 0	12. 5	7		10	
July	12. 1	29 .0	10		34	
August	9. 1	15. 8	4		19	
September	9. 1	8. 6	4		8	
October	6. 1	6. 6	6		1	
November	9. 1	2 .0	3		0	
December	12.0	2. 6	1		2	
Total	100.0	100.0	52		87	1
Month unknown			0		2	1
Grand total			52		89	2

*Habitual users of supply.
†School children.
‡Visitors.
§ Unclassified

16) S. R. Weibel, _et al_

11

purposes. Data sources included state health departments, medical and engineering literature, newspaper clippings, polls of state sanitary engineers and epidemiologists.

Their findings concluded that on the average, one waterborne outbreak that is known about occurred per month with something over 100 persons becoming ill. Some of the illness was quite severe and about two deaths per year occurred from waterborne outbreaks.

There were 94 outbreaks in private supplies and 34 outbreaks in public supplies. Most of the outbreaks were classified as gastroenteritis; these include 38 outbreaks and 25,800 cases.

Viruses in Water Supply

Dr. Luther L. Terry, U. S. Surgeon General, was quoted in '62 as saying, "We are by no means sure that at least some viruses are not slipping through our present water purification and disinfection processes and entering our water mains". [17]

In reporting on the 13th Water Quality Conference, McDermott[18] summarized the reports on virus and water quality as follows:

> 1. Most speakers agree that there is a growing concern with the problem of viral pollution of water in this country.
>
> 2. It was generally agreed that, because the basic properties of viruses are poorly understood, their transmission by the water route has not been fully appreciated. However, sufficient evidence has been presented to support the thesis that virus can be transmitted by water and result in human infection and disease.
>
> 3. The speakers unanimously agreed that there is insufficient technical data now at hand to document conclusively the idea that virus-free water can be attained with our present wastewater treatment and domestic water supply treatment processes and practices.

---

17) Quoted in Report of Committee on Recreation Use of Sources of Public Water Supply, N. Eng. Water Works Association, Journal New England Water Works Association, 85:1:92 (1971).

18) J. H. McDermott, Ibid

4. Thus, immediate attention should be given to additional studies and investigations in the following subject areas: viral detection and enumeration methodology, basic properties of enteric virus, transmission through the aquatic environment, viral disease of man and associated epidemiological studies, and unit process research and development.

In a discussion of viral disease and epidemiology McDermott reports that we can reasonably be assured that the transmission of human enteric diseases are largely dependent upon water as the vehicle.

Metcalf and Melnick presented papers at the conference indicating that one strain of virus may produce illnesses with widely variable incubation periods and manifestations, e.g., a coxsackievirus may produce meningitis in one person, myocarditis in another, and diarrhea in a third, etc. [19]

S. F. B. Poynter[20]

S. F. B. Poynter, Senior Bacteriologist, Metropolitan Water Board, addresses this same area of concern: Infection may be brought about by ingestion of virus contaminated food or water or by downward spread from the respiratory tract. The prolonged excretion of the virus in the faeces which follows infection causes faecal contamination to be a major factor in their spread, and probably most enterovirus epidemics are started by symptomless excreters.

Poynter also specifically indicated that Adenoviruses are aetiologically associated with upper respiratory tract infections.

James W. Mosley[21]

Mosley attempted to classify how often water is responsible for transmission of viruses responsible for infectious disease. He reported that infectious

19) T. G. Metcalf, "Biologic Parameters in Water Transmission of Viruses;" J. L. Melnick, "Detection of Virus Spread by the Water Route," Proceedings of the 13th Water Quality Conference, University of Illinois at Urbana-Champaign (1971).

20) S. F. B. Poynter, "The Problem of Viruses in Water", Water Treatment and Examination, 17:187 (1968).

21) Mosley, James, Ibid.

hepatitis is the only disease caused by an agent having the characteristics of a virus, for which evidence of waterborne transmission has been accepted by all workers in the field. Nonetheless, he considers relatively neglected the entities known as viral gastroenteritis and diarrhea, although numerous waterborne epidemics are documented.

He points out that the limitations of using the term "viral" to apply to gastroenteritis or diarrhea are that it provides a wastebasket for those episodes due to agents other than viruses not yet identified as pathogenic, as well as for those instances in which the search for known pathogens failed because of delay in the investigation or technical inadequacies in the laboratory.

> "Despite these limitations, experimental studies have indicated that forms of gastroenteritis and diarrhea transmissible from person-to-person are due to agents with characteristics of viruses (Gordon et al, 1949; Jordan et al, 1953; Reimann, 1963). It seems plausible, therefore, that under some circumstances such agents could be waterborne. In addition, it has been possible to associate epidemics of gastroenteritis and diarrheal disease in newborns with infection by enteroviruses (Eichenwald et al, 1,958; Lepine et al, 1960). The extent to which enteroviruses are responsible for these syndromes in older children and adults is less clear. Echovirus 11, however, has been reported to have caused gastroenteritis in adults working with this agent in the laboratory (Cramblett et al, 1962; Klein et al, 1960). No association of enteroviruses with waterborne episodes has been achieved, but it is uncertain whether any serious attempts have been made to do so."

Finally, Mosley indicates that evidence for waterborne transmission of viral disease has been based primarily upon epidemics in which cases occurred within a sufficiently short period of time to make person-to-person transmission an unlikely explanation.

There is "feeling" at this time that the role of water transmission of viruses is, therefore, especially related to municipal systems using surface water.

O. C. Liu*

In a personal communication, Dr. Liu addressed the question of dollar costs associated with occurrence of viral illnesses. He believes that from available epidemiologic data, virus is perhaps the most important water pollutant which could produce harmful effects in man.

Pointing out that water may be polluted by biologic, chemical and radioactive pollutants, Liu suggests that by eliminating these pollutants, human health may be benefitted and these benefits may be converted into dollar values. Health benefits are enumerated as either dollars saved for medical expenses as if the patient had not been sick and/or the dollars which could be earned in one's lifetime if the patient dies.

Liu groups diseases produced by enteric viruses into two major categories:

      a) Acute clinical illness
      b) Health effects from subclinical infections

The subclinical infections are of primary concern to us in this study. Liu reports that subclinical infections of enteric viruses in man, representing 100-1000 times the incidence of clinical infections, are far from harmless. Certain numbers of enteric viruses have been associated with serious delayed effects.

He states that the whole group of enteric viruses may potentially be transmitted by water through: a) drinking and culinary use; b) recreation; c) agriculture; and d) food.

---

   * - O. C. Liu, M. D. Personal Communication, Office of Research and Monitoring, Environmental Protection Agency, Narragansett, Rhode Island (1972).

With the present knowledge, therefore, he estimates that health benefits could be of two extremes:

(a) One extreme is represented by use of the published data on waterborne disease incidence as a basis for the health benefit estimate. Practically, these include only two diseases: infectious hepatitis. and gastroenteritis or diarrheal disease. But he personally feels that the reported figures (Mosley, Wiebel, et al, etc.), probably represent only a fraction of those which actually occurred. "If we estimate health benefits according to the above, the total value will be too low to be realistic or true."

(b) The other extreme is to estimate the benefits by assuming that eradication of all enteric virus diseases incidence may be achieved if the water pollution from virus is under control. Liu feels there is circumstantial evidence to indicate that the latter might be the case. "Intensive research, however, is badly needed in this area in order to definitely identify this problem. The long and insidious route of transmission by water is not at all understood at present. Water, however, may serve as an essential link in the chain of events in perpetuating these viruses. Once this link is broken, the virus finally may stop to exist among human hosts thereby eradicating these diseases."

Liu gives his estimated dollar value which may be saved per annum, if indeed eradication of enteric virus diseases in this country can be accomplished by control of water pollution. For acute gastroenteritis and diarrheal diseases (2,000,000 work days lost/year at $30/day) the rough estimated figure is $60 million.

Health & Mineral Characteristics of Water Supplies

Neri, Hewitt and Schreiber[22]

In a recent review for the Journal of Epidemiology, Neri, et al present a synopsis of findings on literature investigating relationships between health and mineral characteristics of local water supplies. They remark on the general impression resulting from the literature reviewed, (53 journal articles and papers) "that some 'water factor' does exist, but no agreement is in sight concerning the identity of this factor, its likely mode of action or even the pathological effects produced by it". [23]

---

22) Luciano C. Neri, David Hewitt and George B. Schreiber, "Can Epidemiology Elucidate the Water Story?" Journal of Epidemiology 99:2, February 1974, 75-88.

23) A comprehensive review of the literature is contained in Masironi, R., Miesch, A. T., Crawford, M. D., et al, "Geochemical Environments, Trace Elements & Cardiovascular Diseases," (forthcoming). Another aspect of water mineralization damages can be found in Tihansky, Dennis, "Economic Damages from Residential Use of Mineralized Water Supply", Water Resources Research, April 1974.

An overview of the findings from this literature review indicates that water quality relates to cardiovascular disease, or some subdivision of it. But as Neri, et al point out, the support for this finding is not yet even very strong. In fact, "it has been progressively weakened by the admission of bronchitis 'effect', [24] by the suggestion that there is also an infant mortality 'effect', [25] and by the estimate based on Canadian statistics that more than half the excess mortality in soft water areas is certified to noncardiovascular cause of death. [26]

In their review, Neri, et al emphasize the need for an increase in epidemiological studies conducted a systematic fashion, with due regard to the need for specifying and testing explicit hypotheses. They recommend studies which utilize a method such as partial correlation analysis. Their stress on more rigorous analyses in the field of epidemiology is illustrative of a weakness inherent in research - no matter what the discipline - which accents descriptive and anecdotal criteria in the achievement of results.

Coliform and Pathogenic Organisms

T. Virarghavan[27] indicates that in studies of water quality, "The coliform organisms, because of their large numbers in fecal matter, offer a far more satisfactory approach to the detection of fecal contamination than do the pathogenic organisms. A search for pathogenic organisms is very difficult and time consuming and may prove fruitless when fecal contamination is present but specific pathogens are absent". He further points out that the water utility profession relies on coliform examination to determine whether water is free from disease producing microorganisms.

---

24) Crawford, M. D., Garner, M. J., Morris, J. N. "Mortality & Hardness of Local Water Supplies," Lancet 1:827-831, 1968.

25) Lowe, C. R., Roberts, C. J., Lloyd S., "Malformations of Central Nervous System & Softness of Local Water Supplies," British Medical Journal, 2:357-361, 1971; and Crawford, M. D., Gardner, M. J., Sedgwick, P. A., "Infant Mortality and Hardness of Local Water Supplies", Lancet, 1:988-992, 1972.

26) Neri, L. C., Mandel, J. S., Hewitt, D., "Relation Between Mortality & Water Hardness in Canada", Lancet 1:931-934, 1972.

27) T. Viraraghavan, "Water Quality & Health" Journal of the American Water Works Association (October 1973) pp. 647-650.

Virarghavan makes the statement that although pathogenic organisms may survive water treatment that apparently removes all coliform bacteria, the evidence available from outbreaks of waterborne disease from treated public water supplies shows that the coliform examination has been a satisfactory measure of the microbial safety of water.

In this same regard, Krishnaswami[28] states, "There is no evidence that bacterial, enteroviral and other microbial diseases are frequently transmitted to man by water that meets the relatively stringent bacterial, disinfection, and chlorine residual standards for drinking water".

Bathing and Recreational Water

It is of interest to note that while uniformity exists for bacteriological standards of drinking water in the United States, most of the states have a variety of bacteriological standards for bathing water.

The most pertinent study completed in the bathing water quality and health area was carried out by Stevenson prior to 1952.[29] The goals of the research were to determine what frequency of swimming might be expected from the population groups chosen; to determine what relative increase in illness incidence might be expected in the cleanest waters; and to determine what differences in illness incidence might be expected from swimming in water containing various degrees of bacterial pollution.

The studies were conducted over two months with families living near 1) a great lake; 2) an inland river and a fresh water recirculating pool, and 3) a tidal water. Stevenson's overall findings were the following:

a) An appreciably higher overall illness incidence may be expected in the swimming group over that in the non-swimming group.

b) When the total illnesses among swimmers and non-swimmers were compared, there appeared to be no significant correlation between the illness incidence and the quality of water in the areas studied.

c) Some 100 percent higher illness in the group under 10 years of age was shown than for those over 10 years of age.

d) Among swimmers, eye, ear, nose and throat ailments represented more than 50 percent of the illnesses reported; gastrointestinal disturbances about 20 percent and skin irritations the remainder.

28) Krishnaswami, p. 2262

29) Stevenson, A. H., "Studies of Bathing Water Quality & Health", American Journal of Public Health, 43:5, 1953, 529-538.

e) Specific correlation between illness incidence and bathing
in waters of a particular bacterial quality was observed in
two instances. Illness frequency was significantly higher
among swimmers when the average total coliform density
was up to 2700 per MPN rather than below 43 percent MPN.

Stevenson was particularly cautious about the latter finding as only three days
were on each side of the comparison; plus, thorough analysis of the effect
that weather might have had on the result was not permitted.

McKee indicated findings which showed that sewage-polluted sea water
carried only negligible risk to health unless the water was "highly" polluted. [30]

What remains unknown is the generalizability of these findings across time
and place. Research in water recreation and health has received particu-
larly little attention, due to the difficulties inherent in carrying out a valid
study (i.e., data collection problems, costs involved in a study of statis-
tically significant proportions, difficulties with data multi-collinearity,
etc.). Nonetheless, the findings described above, as well as other studies
which reveal rising illness in the summer months, indicate the need for
program emphasis in this research area.

Summertime Gastrointestinal Illness

Petersen and Hines carried out a study on the relationships between summertime
gastrointestinal illness and the sanitary quality of water supplies in six rocky
mountain communities. [31] The research performed is notable for its employment
of rigorous statistical analyses. Additionally, Petersen and Hines identify their
analytical methods and data sources.

Their query examines existence of a correlation between unreported gastro-
intestinal illness and the populations using the water supplies. Approximately
100 families in six Rocky Mountain communities were surveyed. No schools
were in session during the period.

Variables included: family experience with gastrointestinal illness during June,
July and August of 1957. (Gastrointestinal illness was defined as diarrhea with
or without nausea and vomiting, persisting for at least one day); and bacterio-
logical water analysis for each town over the past 3 years.

Illness experienced was obtained by means of survey on a house-to-house basis
by 2 investigators, using a standardized form, directed to one responsible
household person, usually the mother. Data were recorded on McBee cards
specifically designed for gastrointestinal illness. Table compilations were
made up of the observations.

30) McKee, J. C. and Wolf, H. W. "Water Quality Criteria", California
State Water Quality Board Publication No. 3-A, 1963.

31) "The Relation of Summertime Gastrointestinal Illness to the Sanitary
Quality of the Water Supplies in Six Rocky Mountain Communities",
American Journal of Hygiene, Vol. 71:314-320, November 1959.

Bacteriological water analysis was taken from the office records of the State Health Department. Numerical values were assigned based on a demonstrated consistency of quality over a three year period.

Data manipulations were conducted by means of simple chi square tests of epidemiological factors associated with the illnesses showing significant differences between the group of communities with water of low sanitary quality and those in the group having high sanitary quality.

Their findings concluded that no patterns could be found from analysis of chronological arrangement of cases of illness in each community by date of onset, which could be interpreted as being the result of a common source, single exposure type of infection. Only 20 of the 206 cases showed a history of eating away from home during the week preceding illness. No correlation was found between incidence of illness and particular brands of milk.

No correlation was found to environmental sanitation deficiencies, because, of 570 premises surveyed, all had running water inside the house, and only three had sewage disposal of a type other than an indoor, water-flush system.

Communities with water supplies of low sanitary quality were found to have an overall attack rate of 13.9 percent, while communities with water of high sanitary quality had an overall attack rate of 8.8 percent. This correlation. between water quality and attack rate was found consistently when computed separately for each community.

The attack rate for males was higher than the rate for females in both groups of communities with no significant difference. A concentration of illness among the young appeared to be greater in communities with water supplies of high sanitary quality. Attack rates computed by family size indicated no consistent gradation. The percentage of cases occurring in July and August indicated a definite increase as the summer progressed, for both groups of communities.

Interestingly, an overall attack rate among people who had lived in the communities with low sanitary quality water supplies for less than two years, was twice that of the people who had lived in the communities more than two years.

Benefits from Reduction of Waterborne Disease Outbreaks

Jack Lackner did the only cost-benefit study we found at all attributable to water quality effects on some low level illness. [32] He queried the extent to which benefits from anticipated reduction in waterborne disease outbreaks and health damage can be predicted! His point of reference was the Safe Drinking Water Act of 1973.

---

32) "Safe Drinking Water Act of 1973: Estimated Benefits and Costs:, EPA Office of Planning and Evaluation.

He examined disease data in the United States, 1961-1970; flourida-tion practices in the United States (1967 capital investment data con-sidered in relation to 160 million people served); and, future water supply system upgrading (costs for 1600 supplies).

Data and sources were: 1) Disease incidence from Morbidity and Mortality, annual supplement, Summary 1970, USDHEW, Center for Disease Control; 2) Bacteriological data, effectiveness of flourida-tion, and lead excess data from the "Community Water Supply Study"; 3) Capital investment data in public water supply facilities from 1967 Department of Commerce data and; 4) Department of Commerce and the American Water Works Association estimates on future costs of upgrading water supply systems to comply to 1962 PHS Drinking Water Standards.

Data manipulations included the following: 1) Judgmental estimates were made by Public Health Scientists that 1/3 or approximately 245,000 of reported disease cases are in fact waterborne. An assump-tion that the recording of cases of waterborne disease is about 10 per-cent of actual number of cases of waterborne disease, increased the number to 2,450,000 in the 10 year period. 2) Estimation of in-come at $25 per day and hospitalization at $200 per day. 3) Assump-tion of 1,000,000 cases of drinking water gastroenteritis per year at comparable costs of $50 per case. 4) Correction of poor flouridation using "Community Water Supply Study" indicator of 38 percent of the 86,136,000 people receiving ineffective levels of flouridation as it pertains to dental disease. 5) Capital investment data in public water supply facilities considered in relation to the 160 million people served at $312 per capita. 6) Assumption of three chlorinators needed per system for $24,000 systems at a cost of $3,000 each. 7) Community water supply figures indicating that 1,000 supplies need defluoridation equipment, and 600 supplies need correction of failure to meet constituent standard of lead.

Lackner's findings included a total for estimated annual benefits of $810,000,000. This figure was derived from the $306 million attribut-able to reductions in communicable diseases; $50 million attributable to reductions in the incidence of gastroenteritis; $454 million attribut-able to reductions in dental decay.

DOC and AWWA data estimated that $6 billion should be spent at a rate of $500 million per year over a 15 year period for improvement of water supply facilities. The yearly cost of program to correct inadequate design, construction, operation and maintenance of water supply facili-ties, was estimated at $2.5 billion. The direct Federal cost of imple-menting legislation on water quality was estimated at $27 million in FY '74, increasing to $63 million in FY '78; estimated cost to states was $50 million each year; estimated costs to communities was $165 million annually. The additional future costs required to upgrade all water supply systems was estimated at $6 billion, to be spent at the rate of $500 million per year.

Summary

Overall, the Petersen and Hines morbidity study revealed a significant relationship between the sanitary quality of a community's water supply and the incidence of gastrointestinal illness in that community. Communities served by contaminated water suffered an attack rate of 14 percent during the summer months as opposed to the 9 percent attack rate for people using water supplies showing no contamination.

The latter finding corresponds to the Weibel, et al, peaking of waterborne disease outbreaks due to private supplies, the summer months.

Lackner's analysis resulted from studies showing that 30 million people, served by 8,000 community water supply systems, were receiving inadequate flouridation in their water for maximum benefit to their dental health. Furthermore, 64.4 million people, served by approximately 16,000 systems, were receiving water containing dentally insignificant flouride ion levels. His cost benefit study gives rough estimates of monetary and health savings derivative from effective flouridation levels, which are calculated to be in the millions. No other researcher to our knowledge has published findings comparably achieved.

Of the literature reviewed relevant to water quality and low level illness, the above two studies were the most analytically rigorous, and empirically definitive.

More precise analyses have been undertaken in the area of air pollution and health, several of which are quantifiably verifiable. Particularly notable is work done by Ridker [33] on morbidity, mortality and respiratory disease, and the subsequent studies by Lave and Seskin [34] which expanded the diseases covered by Ridker. Ridker estimates that 18 to 20 percent of the approximately $2 billion in national health costs results from air pollution. Lave and Seskin, in a conclusion which they consider to be conservative, estimate that air pollution damage amounts to 4.5 percent of all the economic damage associated with morbidity and mortality.

Similar findings are being reported which specifically indicate relationships between air pollution and serious illness. For example, a recent report by Henderson, Menck and Casagrande of the University of Southern California indicates that residents of South Central Los Angeles, who have for years lived near oil refineries and chemical plants, are more likely to die of lung cancer than people who reside elsewhere.

33) Ridker, Ronald G., Economic Costs of Air Pollution.
    New York, Frederick Praeger, 1967.
34) Lave, L. B. and E. P. Seskin. "Air Pollution & Human
    Health", Science 169 (3947): 723-733, August 21, 1970.

In a study which specifically addressed air pollution vis-a-vis low level illness, residence data again proved of consequence: Chapman, et al, queried to what extent air pollution exposures related to the prevalence of chronic respiratory disease. [35] Using data from New York, findings suggested that moving from a polluted area to a clean one can promote substantial improvement in symptoms.

Other evidence from the same study suggested that sulphur oxide pollution exerts an effect on chronic respiratory disease which may rival the effect of cigarette smoking, particularly in areas where high exposures to sulphur oxide are coupled with suspended particulates.

A final air pollution/low level illness study by Thompson, et all, investigated the effects of air pollution on the common cold. [36] Using multiple regression it was found that statistical analyses differed markedly by season of the year and by whether incidence or common cold was being examined. Meteorologic variables used appeared to be more related to the common cold rates than the pollutant variables, but examination of the data failed to "explain" the reasons underlying this finding.

Overview of Findings

While highly definitive findings examining water pollution relationships to low level illness are still in the future, of the studies conducted to date, the evidence points to probable significant relationships. An overview of findings from the literature reviewed can be summarized as follows:

• The role of water as a causal factor in occurrence of low level illness has not received serious consideration by scientists. This results partially from the fact that in both water and air pollution studies, the role of environmental pollutants in mortality and chronic disease is difficult to quantify because so many other determinants of death and disease cannot be adequately measured. A simple relationship of causal specificity does not always exist.

• Studies of waterborne illness covering the years 1920-1936, and 1946-1960 have classified over 80 percent of the cases as gastroenteritis or diarrheal illness. While not specific diseases because no etiologic agent can be identified, these categories of illness account for the vast majority of American illness due to waterborne sources.

35) Chapman, Robert S. et al, "Chronic Respiratory Disease Symptom Frequency in Late Adolescence, In Parents of School Children and in Elderly Persons", Presented at the American Medical Association's Air Pollution Medical Research Conference, Chicago, Ill., October 1972.

36) Thompson, et al, "Health and the Urban Environment: Air Pollution, Weather and the Common Cold", Journal of Public Health, Vol. 60, No. 4 (1970), 731-739.

- Additional studies are needed in the area of viral detection, enumeration, transmission, and elimination because their basic properties are poorly understood.

  There is sufficient evidence that virus can be transmitted by water and result in human infection. However, intensified research has yet to prove conclusively whether bacterial, enteroviral and other microbial diseases are frequently transmitted to man by water that meets the bacterial, disinfection and chlorine residual standards for drinking water.

  Past evidence for waterborne transmission of viral disease has been based primarily upon epidemics in which cases occurred within a sufficiently short period of time to make person-to-person transmission an unlikely explanation.

  There is evidence that viruses may produce widely variable incubation periods and manifestations in different individuals.

- Some sources indicate that the role of water transmission of viruses is especially related to municipal systems using surface water.

  Although present water purification technology can handle all significant bacterial problems, in practice, available technology is not always applied. Many water systems are overage, substandard and unable to meet peak demands.

  Some morbidity studies reveal a significant relationship between the sanitary quality of a community's water supply and the incidence of gastrointestinal illness.

- Research is needed in the area of bathing and recreational water to study indications that swimmers suffer from eye, ear, nose, throat ailments, and gastrointestinal disturbances more often than non-swimmers.

- Two studies have shown a peaking of waterborne illness in the summer months, particularly as this season progresses.

- Residence data in recent studies indicate that longevity in a polluted area contributes to symptoms of serious diseases. However, an overall illness attack rate among people who have lived in communities for shorter periods of time can be observed as being higher than for people who have lived in communities for longer periods (for example, more than two years).

- One estimate states that $60 million can be saved per annum through eradication of gastroenteritis and diarrheal disease from water pollution.

SECTION V

RESEARCH DESIGN

The Problem

The driving question of this research project developed from the CWSS study of drinking water around the nation. If, indeed, drinking water is not uniformly good and water treatment standards are not consistently followed, does this have any health implications with economic effects?

There is an image that our drinking water is pure and that there are no serious health problems resulting from it. It is now apparent that the first half of this image is shattered. Yet there are large numbers of epidemologists who would still argue that there are no serious health effects. They will point out that there are very few known outbreaks of epidemics which can, with certainty, be traced to drinking water. They will argue that we produce anti-bodies to protect us from most bacteria so that even if our drinking water is poor, we will develop immunity to most diseases it could produce. Mexico is indicated as a prime example of a country with contaminated drinking water which presumably does not affect the health of the residents.

To a very large extent, these critics are correct. Indeed there have been isolated known water-related epidemics. We do develop effective anti-bodies and the water probably could get considerably worse before epidemics of major serious diseases occur.

Despite this, the critics may be missing a very important point. People do not have to contract a major disease to get sick. Unreported sub-clinical illness may not be fatal, but it is unpleasant and causes lost work and occasional medical costs. And secondly, the absence of epidemics does not rule out the possibility that the drinking water may carry micro-organisms which do, in fact, result in illness at a sub-epidemic level.

As we have said above, the overwhelming majority of health related absenteeism comes not from the "serious" diseases such as cancer, heart attacks, etc. but from sub-clinical illnesses such as flu, upset stomachs, and colds.

We believe that if our drinking water has any appreciable economic effect on the community, it will be through low level illness. Variations from community to community in this type of illness could easily be significant yet go unnoticed.

Data

Given the above general hypotheses, it is necessary for us to operationalize the concepts and to postulate a set of more specific data-based hypotheses. In doing this it is useful to describe the data which we would ideally want in order to best test our general hypotheses (not that we could get it, because we always have to make some concessions to the reality of availability.) Knowing what we would want, however, enables us to better understand the characteristics of the results which we obtain with real data. Ideally, to perform a study with these goals, we would like a good measure of adult illness which would reasonably be related to drinking water. Ideally again, this would include only those sub-clinical illnesses whose virus or bacteria can be transmitted by drinking water. We would want a good measure of water quality including average coliform counts for a year, and the maximum coliform count for a period perhaps as long as a month.

Additionally, we would want measures of community socio-economic characteristics such as average income, population density, racial distribution, and housing characteristics. Finally, we would want to have average and extreme values for particulate counts and sulfur oxides.

Absenteeism Data

Obtaining direct measures of adult sub-clinical illness is virtually impossible without extensive surveying. The problem is that there is no clear measure of low level illness by community for those communities which we also have water quality data. The National Health Survey is an excellent measure of illness of all types, but it is a nationwide sample which cannot reasonably be broken down to the small communities in which we are interested.

People generally do not report low-level illnesses unless they are particularly severe or prolonged. Thus, doctor's records, hospital records and public health records are quite inadequate for our purposes.

Since we are primarily concerned with finding the economic costs associated with inadequate drinking water, it is reasonable to consider only those incidents of illness which resulted in absenteeism.

Although this simplified the task of measuring community health to some extent, it clearly was not without problems. Direct measures of adult absenteeism are difficult to gather and are of questionable value even where available. As we shall demonstrate, indirect measures appear to be superior. First, however, let us deal with the problem of direct measures.

Any community of 5,000 or more is large enough to have a reasonably diverse set of business enterprises. There may be, for example, one or two medium-sized factories, a few small manufacturing concerns, an array of retail establishments, a small number of construction firms, some government officials, and perhaps some mining or agriculture-related businesses.

Records of absenteeism are not kept uniformly. Some companies have detailed records, other are very incomplete. The variation occurs along at least two dimensions: size of business and style of management. Normally large concerns keep better records than small businesses as a matter of necessity. Furthermore, just as a matter of personal style, some managers simply keep better books than do others. This variation is unknown, unmeasurable, and not necessarily random. Thus, even given some acceptable sampling of employees in a community, it is questionable whether the measure of absenteeism would even reflect the number of days for which work was missed. Even if the actual number of days missed were accurately estimated, there are a number of intervening variables which would make the interpretation of these data of questionable value. Occupation-related illness, variations in paid sick leave, insurance policies, commuters who do not live in the cities where they work, etc., all pose problems which would be extremely cumbersome to overcome.

Given these problems, we chose to select a different measure of absenteeism which, although not perfect, has fewer difficulties: elementary school absenteeism. This measure has a number of characteristics which make it a reasonably good measure of water-related absenteeism.

It is such a large proportion of the poulation of elementary school age children, that the measure can be considered to be the population reflecting the socio-cultural and economic variations of the communities within which they reside.

Unfortunately, the purity of population is often somewhat diminished by the presence of students bussed in from smaller rural communities. Obviously the water which these "outsiders" drink is different from that of the local children. Although complete compensation for the problem cannot be achieved given the resources available to us, there are three factors which tend to relieve the problem. First, the selection of elementary schools (rather than middle or high schools) reduces the problem considerably. School centralization is far less frequent for elementary schools than for higher grade levels. During the course of this contract we were pleased (for scientific reasons) to find that large numbers of small communities will have their own very small elementary schools rather than selecting to send their children to larger cities a few miles away. Secondly, although when they are at home, children from outside a school district are exposed to drinking water of possibly different quality than was measured by the CWSS study, when they are at school, they will drink the same water as other students.

Finally, in one state (Vermont), the records of children who are bussed in from outside the school district are kept separate from those within the school district. Thus, in this state, the records available to us reflected primarily only those children who actually lived in the school district. Since the CWSS study recorded all public water systems in each town visited, this means that for Vermont, we have a high confidence that the student population is nearly entirely drawn from within the water system boundaries. [37]

A second advantage of elementary school data over its adult counterpart is that occupational-related absentees simply are not there. While variations in school health programs may produce some variations in absenteeism, the problem cannot compare with the variations that surround different occupations.

A third advantage is that there is no equivalent to paid sick-leave days. The adage that "sick days taken expands to meet sick days allotted" is not applicable.

A fourth advantage of elementary school data (over higher grades) is that unexcused absence is not as prevalent. Certainly it exists, but the overwhelming number of educators consulted expressed the opinion that unexcused absenteeism is more severe in the high schools than the elementary schools.

---

37) Since we were able to analyze data at the state level, it will be seen later on in this report that we used Vermont as a check for those areas where the data were aggregated only by school district.

Finally, although there are a number of socio-economic factors underlying much of the absenteeism including both excused and unexcused absences these are systematic and analyzable. The economic state of the community, housing characteristics, etc., can all be engaged to account for some of the variation unrelated to water.

Probably the most widely raised question about the use of school absenteeism data has been the fact that we do not know what illnesses caused children to miss school. The child with the broken arm is recorded exactly the same as is the child with gastroenteritis. The former obviously is not water related - the latter may well be. How, our critics will argue, can we possibly make statements about the health effects of poor drinking water when we have no idea what illnesses the children had? The problem is resolvable given one assumption: that controlling for all other systematic influences on health, non-water related illnesses occur randomly. That is, once we account for all systematic effects from economic well being, air pollution, urbanization, etc., the remaining illness is distributed randomly except for water-related illness. If the assumption is true, we have a classic problem of data with random noise. Its effect (Rummel 1965) will serve only to reduce the "real" correlation. In summary to our critics on this point, we can say that if they accept the assumption of randomness, then results cannot possibly be inflated and probably are reduced from what we would expect if we were able to control for that randomness. [38]

Data Collection Procedures

Elementary school attendance data was collected for school year 1969-70 for all communities having at least one elementary school and having been studied in the 1969 CWSS study.

Since both the 1968-69 and 1969-70 school years overlap the calendar year 1969, we had the option of choosing either for our data base. Although the 1968-69 school year covers more months of 1969 than does the 1969-70 school year, we chose the latter because of the wide-spread Hong Kong flu epidemic from November 1968 through March of 1969. We believe that this could have introduced additional uncontrolled variance. At its peak, the epidemic was afflicting 10 million people per week. [39]

In the collection.of our data, we found that central records are normally available in the state capitols or county seats. We were courteously

---

38) Although the correlations will be low, the regression coefficients will still be unbiased.
39) Wilder, Charles S., "Acute Conditions Incidence and Associated Disability." U.S. Department of Health, Education and Welfare, Public Health Service, 1972.

given access to these records by state and county officials. In nearly every state at least a small number of individual schools had to be contacted because the state did not have their communities recorded separately. Where central records were available only the smallest of the communities had to be contacted separately. In five states (Kansas, Missouri, Louisiana and California) attendance records were not kept in a central location and all schools had to be contacted individually.

Water Quality Data

We used two direct measures of drinking water quality and 18 indirect measures covering various aspects of water treatment facilities and their direct operators. These data, shown in Table 3 below were taken from the Community Water Supply Study (CWSS) performed in 1969 by the Public Health Service (PHS). [40]

Unfortunately, the CWSS did not take adequate precautions to insure against the possibility of stochastic variation in drinking water quality. The major portion of the data, in fact, was collected at one point in time. This would be adequate if drinking water quality was relatively stable across time. However, it is a widely held belief that for many reasons, the quality of U.S. drinking water can vary considerably across time. Since we have only a one time measure of water quality for any particular tap, it is necessary to assume that water quality is stable across time if the measure is to be acceptable as a linear function of an average across an entire year. In most instances, this is a dangerous assumption. Therefore, the results which will be discussed in this report must be evaluated with cognizance for the fact that the water quality data were not collected in a rigorous time series.

---

40) McCabe, op. cit.

Table 3 [41)]

Total Coliform
Total Plate Count
Range of least and most monthly samples
Is the lab certified?
Frequency of sampling of finished water
Common walls between finished and lesser quality water?
Are there inter-connections to other systems, of known acceptable quality?
Of unknown quality without protection?
Detectable chlorine residual in distinct parts of the distribution system?
Chlorination process interrupted XX times in last 12 months
Highest level of formal education
Length of time on this job
Total years water purification experience
Is operator full time employee?
Salary range
Level of training in water treatment
Level of study in microbiology
Level of study in water chemistry

41) Heavy metals and toxics, etc. were not included in the study. The reasons are that 1) the presence of heavy metals is so infrequent as to make statistical analysis unreliable; and 2) the effects of heavy metals are not likely to show up in school attendance data.

Fortunately, the CWSS did provide us with a number of potential surrogate measures which make intuitive sense, and which we also found to bear some interesting relationships to school absenteeism. These measures are the indicators of plant facilities, procedures and of the operators' capabilities. Logically we expected that poor quality operators and/or facilities would tend to produce poor quality water. Although we have no data which specifically enable us to compare the surrogate with the actual variable,[42] the linkage seems to be a reasonable one. Additionally, the surrogates (salary of operator, education of operator) are stable across time and have some significant theoretical linkages to drinking water quality which will be discussed below.

Air Quality Data

The air quality data was obtained from EPA's National Aerometric Data Bank: Yearly Frequency Distribution. Although SO_2 and particulates are both important measures, only the latter was sufficiently available to permit its usage.

Unfortunately, air quality monitoring stations are not necessarily located in the same communities which were studied in the CWSS study. We, therefore, had to attempt to estimate the quality of the air in the CWSS communities, by using data for the quality of air in nearby communities.

The estimations were made by geographically locating the set of CWSS communities being examined in this study and the most proximate air monitoring stations. Our first rule was that we assume that the air quality is equal to the air quality of the closest monitoring station within 15 miles. If within 15 miles, two or more stations were approximately equidistant from the CWSS community, the one which came closest to being upwind, given the prevailing winds, was selected.

---

42) We cannot compare water quality with the surrogates because we do not trust the direct measure of water quality. Clearly, if we had direct measures of water quality which we believed, the surrogates would not be necessary.

In many instances no air quality stations existed within 15 miles. Indeed, in some areas no measures were to be found within a 50 mile radius. This posed a dilemma because, on one hand, we wanted to retain the option of using all the data -even if questionable - but we also wanted to have the option of rejecting those air quality data associated with stations more than fifteen miles from the city in question. Therefore, we created a dummy variable whose value is 1.0 for cities with air data gathered within 15 miles and -0.0 for those whose data were gathered at a distance greater than 15 miles. This gave us the option of treating the air quality data as "missing" if it were gathered at a distance of over 15 miles.

Socio-Economic Data

The socio-economic characteristics of the communities were obtained from the 1970 census. We used the tapes containing the first count data. From the list of variables available on the first count, we selected the following subset:

1. Total population
2. Percent rural
3. Percent SMSA
4. Percent urban portion of SMSA
5. Percent rural less than 2500
6. Percent in urbanized areas
7. Percent less than 18 years old
8. Percent 18 to 62 years old
9. Percent 62 and older
10. Percent white
11. Percent black
12. Percent of families with a husband and wife
13. Percent other male head of household
14. Percent other female head of household
15. Average home value
16. Average rent
17. Percent units less than 1.0 persons/room
18. Percent units 1.01 - 1.50 persons/room
19. Percent units greater than 1.50 persons/room
20. Percent units with complete kitchen
21. Percent units with complete plumbing
22. Percent without complete plumbing
23. Percent units with shared toilet
24. Percent persons with complete plumbing
25. Percent persons without complete plumbing
26. Percent families with complete plumbing

We believe that this subset of community characteristics are the most likely to have some influence on health. The 26 variables are not all mutually independent. In fact, they fall into six identifiable categories as follows:

Variables	Category
1-6	Urban/Rural
7-11	Age/Ethnic
12-14	Family Characteristics
15-16	Wealth
17-19	Density (persons/room)
20-26	Home Plumbing

Of these six categories, we expected only the last five to enter directly into regressions. Given the results of the CWSS study, we expected that the urban/rural distinction would present particular problems which would require a dichotomization of the sample rather than simply an additional term in the regression equations. The necessity to dichotomize the sample will be discussed with the other hypotheses in the following section.

It should be noted that for each of the remaining five sets, it is only reasonable to use one of the possible variables at a time because they are either statistically or sometimes linearly related to each other. For example, since variables 12 to 14 must equal 100%, any two of necessity define the third. Furthermore, the correlations between any two are expected to be so high as to prohibit the use of more than one because of the problem of multi-collinearity. We chose to include each of the variables in the data set for pragmatic purposes. There is no research or theory which would give us good reason to choose one version of the same variable over another. Rather than arbitrarily exclude the data, we chose to include each version in our initial analysis. We employed a modified stepwise regression by limiting our final variable set to one variable per category. We felt that this relatively mild limitation (one variable per category) made the interpretation more manageable as well as alleviating somewhat a troublesome degree of freedom problem.

Climate

We included the climatic measures of average hours of sunshine, the ratio of temperature/humidity, and inches of precipitation to attempt to estimate the effect of climate on health. These climatic variables were

included because it is widely recognized that weather does influence health. The fact that sub-clinical illness is the greatest during winter months is virtually unquestioned. Additionally, it is normally believed that the effect is most severe in the colder and more humid areas. It should be recognized, however, that in this study we had no time series absentee data to pick up seasonal illness variations and we had only nine geographical regions thus giving us only nine data points for our climate variables. Because of this, our expectations were not particularly high that we would find that sunshine or the temperature/humidity index would produce interesting results.

Hypotheses

If health is to be statistically linked to water quality in a natural experimental setting, it is necessary to us to recognize that, in un-controlled situations, people become ill for a wide variety of reasons. Some of these may have to do with specific illness-producing substances such as air pollution. Others may have to do with inadequate diets or poor health care. Still others may result from living in overcrowded conditions or sub-standard housing. To the extent that these factors systematically influence health, we attempted to account for them in this study. At the most general level, then, we are hypothesizing that, accounting for all other macro-systemic influences on health, the occurrence of sub-clinical illnesses will be the highest in those communities with the worst water quality. More specifically, we are arguing that these factors account for absenteeism roughly according to Figure 1 (next page).

We are arguing that the elements in the boxes act directly on elementary school absenteeism. The elements in the diamonds act on absenteeism through the elements in the boxes. All of these are encompassed in the broader circle which includes those factors which have complex and wide ranging effects on all of the factors and/or their relationships with absenteeism. Because these factors act in concert, we cannot simply pull out water quality and analyze it by itself. Although the other variables are "just" control variables, they are vital to the successful analysis of the absentee data. Let us consider the control variables first.

Health care is one of the most obvious links to absenteeism. Clearly we would expect that improved nutrition, housing, medical attention, would directly influence health. Assuming that health is linked to absenteeism, we can presume that there is a close linkage between health care and absenteeism.

Similarly, cultural factors can be argued to link directly to absenteeism. The ethics of social or cultural groups in all probability influence the motivation to attend school. This may, on one hand, result in simple unexcused absence. On the other hand, some cultural/economic groups

Figure 1
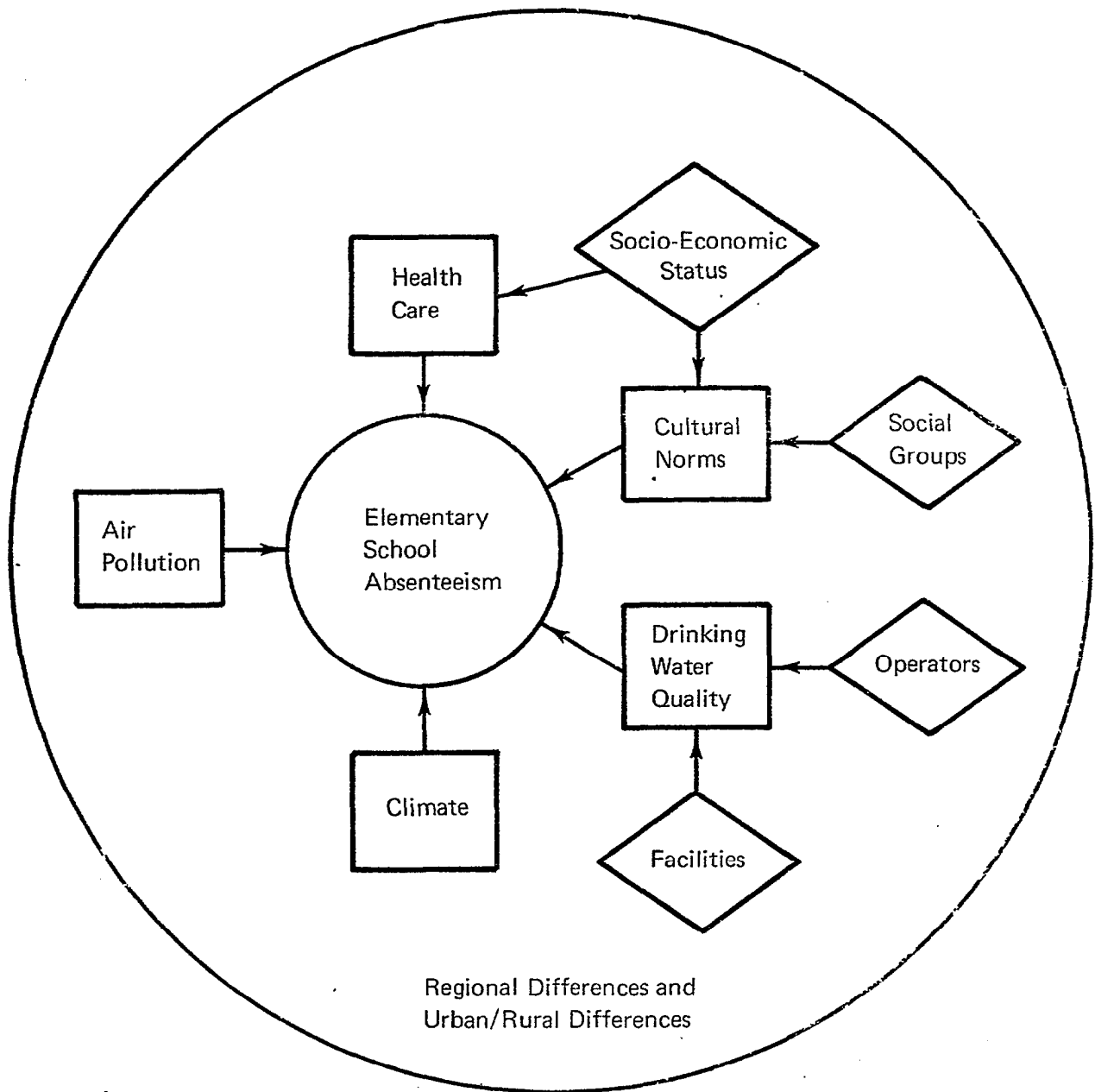
may provide students with encouragement to attend school even when they are ill.

Both health care and cultural factors, we argue, are related to socio-economic status. Normally we expect most of the factors associated with health care to be positively related to socio-economic status (SES). For example, it is common knowledge that the children of low income parents suffer greater physical and mental problems from poor nutrition than do children of more affluent parents. We might also expect that there is some SES level above which there is no general improvement in health care. We also expect that the protestant ethic of the upper and middle classes would have an effect on those cultural norms partially influencing school attendance.

In addition to the economic determinants of cultural norms, it is possible that other cultural descriptors are salient. For example, many people argue that those communities with high percentages of blacks have different behaviors toward education above and beyond the fact that they are normally in lower economic groups. As a point of interest, it should be noted that there was no empirical confirmation of this position.

Air pollution and climate are two of the other factors presumed to directly cause health problems and, therefore, school absenteeism. The exact details of the relationship between air pollution and health are somewhat inconsistent across the numerous studies. Nonetheless, there seems little question that air pollution is related to health in some manner. The relationship between climate and absenteeism is two-fold. On the one hand, as we discussed above, climatic variation influences health. Additionally, adverse weather can close schools, temporarily, obviously increasing absenteeism.

All of these relationships mentioned above are themselves partially and complexly influenced by regional variations and by urban/rural differences. As a simple example, adverse weather often has a more serious effect on the closing of rural schools than on urban schools because of the relatively large miles of highway per person requiring plowing.

Popular wisdom argues that the protestant ethic in rural America when coupled with the allegedly healthier life style should reduce elementary school absenteeism. On the other hand, the frequently lower incomes of rural communities, the more limited access to medical facilities and other factors conceivably could counteract this effect. All

of this is further complicated by the fact that we know for a fact that water treatment facilities in rural areas are inferior to those of urban facilities.[43]

All of these factors basically lead to the fact that these broad complex factors (urban/rural and regional differences) must be handled carefully in the analysis of the data.

Having discussed the control variables, let us now move to the independent variable of interest - drinking water quality. We discussed in the literature review the arguments and analysis supporting the hypothesis that poor quality drinking water can cause illness. That this should logically be the case is not the least bit surprising. What is somewhat surprising is the fact that the water provided through public systems may on a widespread basis, be of low enough quality to produce illness. Given the results of the CWSS and other studies, we have good reason to believe that people do in fact become ill from low quality drinking water. As we see in figure 1, variations in drinking water quality are hypothesized to be influenced by variations in either the quality of the operators or the quality of the facilities or both.

It is widely accepted[44] that the reductions in quality resulting from human error tend to be sporadic rather than steady state. Workers ranging from assembly line operators to air traffic controllers tend to err by making occasional mistakes. They tend not to perform in a constant sub-standard manner. The poorer the worker the more frequent and more severe are the mistakes. Thus, if we were to directly measure the quality of the water produced by poor operators, we would have to have a time series recording the frequency and magnitude of the deviations from the average. Because we do not have data of that quality, we must use measures of operator qualifications as surrogates for water quality.

The CWSS reports that water quality varies regionally and by level of

---

43) McCabe, Op. cit.

44) See for example: 1) Niece, E. H., "Statistical Quality Control and Value Analysis," in Carl Heyel (ed.) The Foreman's Handbook, McGraw Hill, New York, 1967; 2) Levinson, Harry, Executive Stress, Harper and Row, 1970; and 3) Barnes, Ralph, Motion and Time Study: Design and Measurement of Work, Wiley, New York, 1964.

of urbanization. Urban regions tend to have both higher levels of water quality, better trained operators and more sophisticated facilities. Since we suspect that school absenteeism, water quality and the control variables all are influenced by the urban/rural and regional variation, it is apparent from the outset that a generalized model applying to all cases may encounter difficulties. Because of this, we have to recognize the possibility that while it may be possible to develop a general model of the effect of poor quality drinking water, it is likely that we will also need alternative models to deal with the problems of multicollinearity and systematic bias in <u>both</u> the dependent and independent variables.

In addition to multiple models, we will also use dummy variables which themselves include some of the correlated interactions we are trying to simplify. In this study, we added dummy variables representing each of the SMSA's being studied. For example, we had one variable each for Vermont, New York, Cincinnati, etc. If a community were in the region, it received a 1.0, otherwise it was scored a 0.0. This approach enables us to try to include unaccounted systematic regional differences in elementary school attendance. It is important to note that this method is only marginally effective in controlling for any regional bias in the <u>independent</u> variable set.

The primary hypothesis is that one or more versions of the following equation should predict absenteeism.

$$A_{ij} = C_j + \measuredangle_{1j} W_i + \measuredangle_{2j} E_i + \measuredangle_{3j} F_i + \measuredangle_{4j} I_i + \measuredangle_{5j} D_i + \measuredangle_{6j} S_i + \measuredangle_{7j} P_i + \measuredangle_{8j} Cl_r + e_i$$

where
Subscript indicates a water supply within a subset area j

A = elementary school absenteeism
C = constant
W = water quality measured either by plate count of coliforms
    or by quality of facilities
E = one variable from the age/ethnic category
F = one variable from the family characteristics category
I = one variable from the wealth category
D = one variable from the density category
S = one variable from the home sanitation category
P = level of air pollution
Cl = climate
e = the error term.

Frequently, hypotheses such as these are considered proven or disproven simply on the basis of the statistical confidence of the partial correlation or of the regression coefficient. In this analysis, it should be emphasized that we have some definite expectations about the direction of the relationships. These linkages are most important for the pollution variables and particularly for water quality. We would expect both water quality and air quality to be negatively related to absenteeism. Except for some narrow ranges at very high levels of purity, it is not reasonable that higher quality water or air should result in increased illness.[45]

For the other variables, we would expect younger populations to be healthier; non-white population to be less healthy; wealth should be negatively related to absenteeism; density should be positively correlated; and home sanitation should be negatively linked with health. Finally, given the pressures of time and, frequently, income on single parent families, we would expect absenteeism to be higher in single parent families than in the more traditional husband-wife family.

The socio-economic measures would be used in the analysis in an additive manner, but only one variable per group was permitted in the final regression. These were selected using standard stepwise regression.

Analytical Methodology

The primary statistical tool which we employed in the study is stepwise multiple regression. This is a technique which fits well with the general assumptions that health is a result of the simultaneous effects of multiple factors. The methodology provides us with a set of coefficients which indicate the expected magnitude change in the dependent variable as a result of a given change in one of the independent variables.

---

45) At extremely high levels of purity, small increases in bacteria levels may serve as boosters to keep anti-body levels high enough to ward off more serious infectious material. This fact is particularly important for those persons who are a part of a stable population.

In the current study, we are primarily concerned with the health effects of water quality and have taken considerable effort to eliminate any serious problems of multi-collinearity. Since we are less concerned with the individual (as opposed to combined) effects of the control variables, we placed less effort on maintaining statistical independence within this group of variables. We did, of course, take the effort to insure against linear independence of the control variables.

Missing Data

In nearly any statistical analysis using surveys or multiple data files, missing data becomes a problem. There are three logical approaches which can be used to solve the problem only two of which are worth serious consideration. The one logically feasible but impractical solution is to ignore the problem and run the analyses as though the data were present. This alternative inserts zeroes in the missing cells and thereby gives the missing cells arbitrarily incorrect values.

The second alternative is to simply omit any case from the analysis for which there are one or more missing data cells. This alternative has two negative effects which under some circumstances are not particularly difficult to accept. The first is that the sample size is reduced thus reducing the degrees of freedom available for analysis. The second is that if missing cells are systematically distributed, the actual characteristics of the sample are altered by eliminating a non-random subset of cases

The third alternative is to estimate the missing cells given information already in the data set. For example, a data set in which age, income, race, education, and sex were independent variables, a missing "income" cell could be statistically predicted using the remaining four variables. The method is used most frequently in studies having a rather numerous random distribution of missing cells. For example, if each of five variables in a data set had ten percent randomly distributed missing data there would be .5N empty cells. This logically could eliminate as many as one half of the cases.

The difficulty of estimating missing cells is that to do it reasonably successfully creates a conflict of the assumptions of the various techniques. Multiple regressions, as we pointed out above, assumes the independent variables to be independent of each other. The missing data routines, on the other hand, assume that the more highly intercorrelated the independent variables are the more accurate the estimates are. Thus, the closer one moves toward the ideal assumption of regression, the less accurate the

missing data estimates become. There is, admittedly, a gray area in which some researchers believe that the technique is preferable to eliminating large numbers of cases.

In the current analysis, there are two significant types of missing data. The first of these are the census variables which, if any are missing, are generally all missing because we were unable to find the community on the master enumeration district list. Normally this occurs in very small un-incorporated communities which have their own elementary school.[46]

The second primary source of missing data is in the water quality/facility inventory variables. In some instances, the water quality measures were not present. For our sample this was relatively infrequent. More frequently, some of the questions concerning characteristics of the operators or failures went unanswered. These missing cells tended to be scattered although there was a slight tendency for some respondents to refuse to answer several questions.

Given the above configuration of missing data and given that the water quality/facility inventory variables were of primary significance (and, therefore, should not be estimated) we chose to eliminate those cases which had any missing data on the particular variables being analyzed at a specific time. Clearly this would change the sample size as a function of the particular variables being employed in the analysis. For example, consider the following hypothetical data set:

	Var. 1	Var. 2	Var. 3	Var. 4
Case 1	1	1	0	25
Case 2	2	5	0	19
Case 3	3	1	3	
Case 4	2		3	
Case 5	3	1	0	10
Case 6	1	8	6	8
Case 7	1	6	12	8
Case 8	3			32
Case 9	1	1	5	
Case 10	1	2	8	4

------

46) If there were no elementary school, there would be no dependent variable and the case would not even be included in the data set.

If variable 1 were the dependent variable and variables 2, 3, and 4 were independent variables, cases 1, 2, 5, 6, 7, 10 would be examined because they have complete data on all three variables. If, on the other hand, variables 2 and 3 were the independent variables, cases 3 and 9 would be added to the above list because there is data for them for variables 2 and 3 but not 4.

In most circumstances, missing data tended to occur in the smallest communities and those communities with at least one piece of missing data tended to have several missing pieces. The deletion of cases with missing data, therefore, biased the sample in the direction of larger communities relative to the CWSS sample. There is, of course, no assurance that the CWSS sample was itself randomly selected from any larger population.

Case Selection and Aggregation

The CWSS, as mentioned before, covered 969 public water systems across the United States. We were unable to individually examine some of these for three basic reasons:

1) Two or more public systems supplied one community.
2) A public system served a community which did not have an elementary school.
3) The system was in California and the school's records were not easily accessible.

If two or more water systems supplied a single community, a weighted average water quality figure was computed using the number of people served as the weighting figure. For example, consider the following typical hypothetical situation of three systems supplying one medium sized community.

	Population Served	Total Plate Count	P x C
System 1	2000	3000	6×10^6
System 2	4000	4000	16×10^6
System 3	10000	1000	10×10^6
	16000	8000	32×10^6

$$\text{Community means} = \frac{32 \times 10^6}{16 \times 10} = 2000$$

A large number of the systems examined in the CWSS provided water to very small communities, (under 500 population) or to special populations such as hospitals, trailer parks, camping grounds, boarding schools, nursing homes, etc. Of these, all special systems and systems serving those smallest communities with no elementary schools were dropped from the study.

The State of California was one of the states which does not keep centralized absenteeism records. Therefore, individual schools were contacted to collect the absenteeism data for the school year 1969-70. Because absenteeism in California means an unexcused absence, we did not believe that extensive collection efforts were warranted. However, every school district serviced by one of the water systems measured by the CWSS was contacted. If the data for a given school were not easily accessible, we did not request the school's officials to extend themselves to provide it.

The Test

A General Model

Despite all of the potential pitfalls of a general model, we found the results encouraging. The bivariate correlations showed the following variables to be most strongly linked to absenteeism.

Variable	Correlation
Precipitation	.40
Vermont Dummy	-.36
Temperature	.34
New York Dummy	.30
Urbanization	.26
Frequency of Testing of Finished Water	-.24
Louisiana Dummy	.20
Operator Salary	.19

Particular attention should be paid to the incorrect signs for the relationship between temperature, operator salary and absenteeism. We shall find shortly, that these problems disappear when we control for urbanization and regional variation. It is also worth identifying those variables which did not appear in the list of significant bivariate relationships.

Variable	Correlation
Total Plate Count	.09
Home Value	.04
Living Density	.02
Shared Bathroom	.02
Presence of common walls with inferior water	.01
Presence of Chlorine Resident	.03
Air Quality (particulates)	.07

In the general model, none of these variables ever showed significance despite controls for other variations.[47)]

After running the bivariate correlations, we examined the data using stepwise regressions. Because of the complex intercorrelations in the data, we ran the regression several times using different variables. The overall findings from the regressions of all cases is that controlling for urbanization and regional variation, the only consistently significant variable is operator salary. The frequency of testing finished water is consistently related in the proper direction although the relationships are weak. The regression coefficients are in the correct directions in all cases. Operator salary is generally significant at the .10 level. In some instances, however, its significance drops to .20. As we mentioned, the frequency of testing is a weaker variable and is, at best, significant at the .20 level. Interestingly, the product of operator quality and frequency of testing is significant at the .05 level in the following regression equation.

Multiple R = .33 Significance of the Regression = .001 R-Squared = .10

Variable	Partial Correlation	Coefficient	Std. Error	T-Stat.	Significance
Constant		5.7	.57	10.0	.001
Percent in Urbanized Area	.09	.8	.7	1.2	.25
New York Dummy	.20	1.5	.6	2.7	.01
Louisiana Dummy	.26	3.4	1.0	3.5	.001
Fq x Op[48]	-.15	-.08	.04	-2.1	.05

---

47) All of them were significant in at least one regional analysis.

48) An index comprised of frequency of testing finished water times operator salary.

We would argue frequency of testing times the salary of the operator(s) is a powerful index of the operation of a water treatment facility. Whereas, a weakness in one or the other variable may not necessarily mean poor operation, infrequent testing coupled with an underpaid operator almost certainly signifies substandard operation.

Another interesting point is that although temperature and precipitation showed significant correlations in the bivariate analysis, in no instance are they even close to significant in the regressions. Clearly, whatever variance climate was able to explain at the bivariate level was better accounted for by the regional dummy variables in the regressions.

The general model clearly includes a great deal of uncontrolled or semi-controlled variance. The correlations are not particularly strong and a multitude of sins are buried in the regional dummy variables. Nonetheless it is very encouraging to note that variables linked to operation of water treatment facilities are consistently stronger than any of the other competing specific variables such as home value (a surrogate for income), density, cultural factors, etc.

Regional Analyses

In an attempt to gain additional insight into the relationships between drinking water and health, we subdivided the sample into more homogeneous sets. In choosing these subsets it was necessary for us to balance the homogeneity of the subset against the number of cases in the subset. On the one hand, we wanted to have as great homogeneity as possible while also retaining enough cases to permit meaningful analyses. In doing this we finally decided upon four subsets two of which are quite homogeneous, the other two of which are less so by reason of necessity.

The hornogenous subsets are Vermont and the greater New York area excluding New York City. The former includes all of the analyzable CWSS communities in Vermont. The data cover nearly every county in Vermont and systematically exclude only the rural-farm population. We had 82 communities with complete data.

The New York data cover Westchester, Nassau, Suffolk and Rockland counties. The communities studied are generally suburban and smaller urban centers. New York City was excluded because it has an extremely serious unexcused absence problem. After examining the New York City data rather closely, we concluded that uncontrollable variations in unexcused absence made the absentee data hardly meaningful as an indicator of illness. Excluding the New York City data we had 54 cases with complete data.

For the rest of the areas covered by CWSS, we did not have sufficient cases to permit truly regional subsetting. Therefore, we left the areas together and subset them only by level of urbanization. Urban and sub-urban communities were placed in one category. Rural communities as classified by the Bureau of the Census, were placed in the other subset. We had 62 cases with complete data in the urban/suburban subset and 47 cases in the rural subset.

Summarizing, we had four subsets which we subjected to further analysis. They are listed below.

Name	Cases with Complete Data	Comments
(1) Vermont	82	Includes data from most Vermont counties. Generally small communities
(2) New York	54	Four counties surrounding New York. Includes small communities and suburbs with a few medium sized cities.
(3) Six Area Rural	47	Rural communities surrounding Cincinnati, Ohio, Charleston, W. Va., Charleston, S. C., Kansas City, Mo., New Orleans, La., Pueblo, Colorado
(4) Six Area Urban	62	Urban and suburban areas including and surrounding six areas in (3)

The results of the final regressions for the four areas are shown below. The interpretation and discussions will follow the presentation of the equations.

Vermont

N = 82   Significance of regression = .97

Multiple R = .508 R-Squared = .259

Absenteeism = 5.04 - .56S - .33Ic + 16.48D

	Significance	Standard Error of Coefficient
$S^{49)} =$ Salary of operators (index)	.99	1 14
$Ic^{50)} =$ Interconnections with poor quality systems	.85	.21
D = Percentage of living units with more than 1.01 persons/room	.90	8.8

New York[51]

N = 54 Significance of Regression = .99

Multiple R = .614 R-Squared = .377

Absenteeism = 7.02 - 11.4 Env + 34. $25D_2$ - .000033H - $.53R_u$

	Significance	Standard Error of Coefficient
Env = Environment Index	.98	4.4
D_2 = Percentage of living units with greater than 1.5 persons/room	095	16.2
H = Home value	.99	.00001
R_u = Percent rural	.85	.34

---

49) 1 = 1999, 2 = 2000-4999, 3 = 5000-7499, 4 = 7500-9999, 6 = 10000 or
50) 1 = Y, 2 = N
51) Computed by dividing index of operator salary by average particulates. The result is that a high value indicates high environmental quality.

<u>Six Area Grouping - Rural</u>

In the above mentioned six areas, the CWSS studied water systems
in 57 rural communities under 2500 persons which also had an elementary
school.   Of these, 10 had missing data on at least one of the variables
which were finally entered in the regression. The regression results are
summarized below.

N = 47   Significance of Regression = .99

Multiple R = .501   R-Squared = .251

Absenteeism = 13.9 - 1.04F - 1.0S

	Significance	Standard Error of Coefficient
$F^{52)}$ = Frequency of testing of finished water	.99	.32
$S^{53)}$ = Salary of operator (Index)	.90	.54

In this subset we tried to employ the index of frequency of testing
times salary of the operator. The product was significant at the .99
level but the total explanatory power of the regression was smaller than
when we employed the additive model. Because the additive model is
easier to extrapolate and because statistical confidence for the additive
mode: was quite satisfactory we remained with that solution.

---

52) 1 = each month, 2 = each year, 3 = 2 years
     4 = 3 years, 5 = infrequent,   6 = never

53) 1 = < 1999,   2 = 2000-4999,   3 = 5000-7499
     4 = 7500-999, 5 = 10000 or

<u>Six Area Grouping - Urban/Suburban</u>

    N = 62   Significance of Regression = .99

    Multiple R = .79 R-Squared = .62

    Absenteeism = 3.81 + .0006C + .69R + 54.6T + 1.4K

	<u>Significance</u>
C = Total plate count	.99
R = Presence of Chlorine residual	.95
T = Percentage of living units sharing bathroom facilities	.99
K = Kansas City Area Dummy	.90

    In examining the results there are two facts which are initially apparent. The first is that the strengths of the relationships are considerably stronger for the subsets than for the general model. Multiple R's range from a low of .501 in the six area-rural to .79 in the six area-urban/suburban, as compared to multiple R's in the .30-.35 range for the general model. The significance of the water-related variables is also appreciably higher.

    The second major point to observe is the fact that in all regressions except one, operator salary is a significant predictor of absenteeism. In Vermont and the six area-rural areas the relationship is additive with one other measure of plant operation. In New York, operator salary interacts multiplicatively with air quality. It would appear that where there is appreciable air pollution, it lowers resistance to disease such that contaminants in the water will more easily produce illness.

    Although operator salary is a strong predictor of elementary school absenteeism in the national sample and three of the subsets, <u>it is not linked to absenteeism in the six area-urban/suburban grouping</u>. Here we find chlorine residual and total plate count as the primary factors. Because it is only in this isolated subsample that these variables appear, we have to be somewhat skeptical of the confidence we can place in them. Caution is particularly important when we recall that the direct measure of water quality was taken only at one point of time. Thus, the reliability of the independent

variable in this regression equation is in question. Keeping these important caveats in mind, the findings do make sense if we were to hypothesize that for some reason, the contaminants in the water in this subset arose from systematic non-human related problems. For example, if there were some serious problems with the distribution systems, we could argue that the one time measure of total plate count would, in fact, identify the presence of generally contaminated water. Similarly, any facility problems such as these would be more likely to produce a steady level of contaminants than would operator malfunctions. If for some reason, the six area urban/suburban systematically has different problems than all other areas, the results have meaning. Again, however, these results must be taken with caution.

Across the total sets of results, we find that urbanization and geographical location appear to play predominant roles. These, in fact, prompted subsetting in order to control for their associated complex variances. Aside from these factors there is the unmistakable presence of something we might call housing quality. One or more of the variables measuring home value, occupants/room, and the frequency of shared bathrooms occur in three of the four subsets. The importance of these variables is most prominent in the Vermont and New York subsets. As the models covered broader geographic ranges (i.e., six area groupings and the general model) the effect of housing variables tended to wash out. This is not particularly surprising given the geographic differences in property and housing values and types. [54]

Notably, none of the climatic variables enter into any of the equations. It should be noted that weather variables play a strange role in studies which are not time series oriented. Although from one month to the next, weather variation is known to influence absenteeism, this variation can be seriously obscured by high level aggregation. This is particularly important when the communities being studied are geographically proximate to each other. For example, we would not expect the weather in Burlington, Vt. to be much different from Rutland Vt. We know, however, that absentee rates do differ between these two communities. Weather may theoretically explain a good part of the stable base of absenteeism but it cannot reasonably be expected to explain the differences between two geographically proximate communities.

Even in the six area groupings, we have only 3 or 4 apreciably different climates. Here, one results show that the influence of climate is subsumed in the geographical dummy variables. These latter variables apparently incorporate more than climate and are, therefore, the stronger predictors.

---

54) For example, it is difficult to assume that differences in home values between Mt. Kisco, New York and Abita Springs, Louisiana reflect differences in standard of living.

In concluding the interpretation, we would point out that in both the general model and in three of the subsets, operator salary shows a strong and consistent relationship to elementary school absenteeism. There are some differences from region to region but they basically support each other. The one significant deviation occurred in the six area urban/ suburban grouping in which operator salary did not enter at all but was replaced by chlorine residual and total plate count. Given the data collection difficulties, one would have to intrepret this one deviation with considerable caution. If the results of the six area urban/suburban grouping are meaningful, they must be taken to indicate some non-human problem which water.

As support for the major findings concerning operator salary, we should recall that the CWSS reported that 84 percent of the principle operators were earning less than $7500 per year and that 37 percent were earning less than $2000. With such poor salary structures for operators, it is not surprising to find this as a significant element in the production of poor quality water.

The fact is quite significant. While the regression equations broadly predict an improvement for all regions studied of .2 to .5 percent in attendance for each $1000 increase in operator salary, we should not assume that this means that simply by paying the operators more money, we will improve performance by that amount. While there may be some improvement in performance with increased salary, the problem is obviously much deeper than that.

Given 1969 general salaries in the areas we were examining,[55] the majority of operators were in the lower 50% income range. This simply makes recruitment of individuals with college training in chemistry, microbiology and related subjects quite difficult. In fact, the CWSS showed that over 60% of the operators never had even a short course in water treatment. To a larger extent, the public's image of municipal water treatment operators is often much closer to a manual laborer than to a professional with a significant role in community health. According to some people with whom we discussed the problem, the primary role of the operator is to keep the machinery working properly. While this is an important function performed in a water treatment facility, it hardly seems reasonable that that it be the primary role of the principle operator.

Somewhat illustrative of the type of problem encountered is the following anecdote reported by Harris and Breckner: [56]

---

55) The averages fluctuated from area to area between $6000 and $8000 per year per worker.

56) Harris, Robert and Edward Breckner, Consumer Reports, June 1974, p. 436-443.

> The operator in charge of one New England
> community water supply, for example, heard
> somewhere that activiated carbon will remove
> foul tastes and odors. So, when the water
> smells bad, he fills a cloth bag with activated
> carbon powder, ties a rope to it, climbs into a
> rowboat, and tows the bag back and forth across
> the town reservoir. This is about as effective
> as waving a wand over the reservoir.

Obviously to simply increase salaries would not be enough. There would have to be a program to upgrade the operator quality through training programs and, in the long range, improved hiring.

Another important water-related variable is the frequency of testing of finished water This is significant at the .99 level in the rural areas not including New York and Vermont. In the general model it consistently appears with the proper sign although the statistical significance is erratic across different regressions. This finding is also consistent with a major deficiency uncovered by the CWSS. In their report, they discussed the fact that 90 percent of the systems studied failed to meet the bacteriological sur-veillance criteria.

Finally, in Vermont there is a weak relationship between absenteeism and systems with interconnections with water systems of unknown quality without protection. This problem is relatively rare showing up in only six percent of the total systems examined. Interestingly, thirty-five of these were located in Vermont.

Effects

The effects of degraded drinking water can be expressed in a number of ways. The regression equations themselves are the most abstract form of this effect. They essentially provide an estimate of the percentage change in absenteeism, per unit change in the independent water-related variable(s). In the following section we will translate these into more concrete forms moving progressively from the soundest conclusions to the most tenuous. We will initially calculate the magnitude reduction of elementary school absentees in those counties studied resulting from incremental improvements in drinking water quality. [57] The above findings will be expanded to the dollar

---

57)Because of the tenuousness of the six area urban findings, we will not
   extrapolate the costs for this group.

value of lost wages and medical bills saved due to reduced absenteeism/ illness. Finally, making some very bold assumptions, we will extrapolate the total dollar savings to the entire country.

Elementary School Absenteeism - Vermont

Because the CWSS covered nearly every county in Vermont, we believe that the findings apply to all except the rural-farm population of the state. The study applies to a population of 409,000 and an elementary school enrollment of 66,600. The regression coefficient of the variable "salary of operators" is .56, and the coefficient for "interconnections with unprotected systems" is .33. Because the salary variable was coded in a variable interval manner, we can strictly say only that an increase of one point in the salary index will decrease absenteeism by .56 percentage of the total student days. Assuming a 180 day school year, there are 12,000 student days per year and each unit increase in the salary index would result in an estimated 67,100 fewer absences from elementary schools. The standard error of the regression coefficient is .14. Thus, a reduced absenteeism of from 50,000 to 84,000 days is within the plus or minus one standard deviation unit A range of 33,600 to 101,000 is the range defined by plus or minus two standard units.

Since the salary index intervals are roughly $2500 increments, we introduce only a small error by translating these into actual salary values. Doing this we would estimate that, state wide, every $1000 increase in the average salary of the principle operator would reduce total elementary absenteeism by 27,000 days per year.

The interconnections with unprotected systems have a regression coefficient of .33 indicating that for each ten percent of the total systems which eliminate the problems, absenteeism rates will be reduced by .033 percent of the total school days. This translates to an increase of 4000 student days per year. The standard error in the regression coefficient is .21. The comparable range of reduced absenteeism for one standard deviation is 1500 to 5500 per ten per cent increase in eliminating linkages with unprotected systems. The range for the plus or minus two standard units is 0 to 9000.

If we are to draw any inferences about the linkages between drinking water quality and adult absenteeism, we must be able to make some

assumptions about the relationship between adult absenteeism and elementary school absenteeism. Our first assumption is that the portion of elementary school absenteeism explained by drinking water variation is a subset of <u>acute</u> illness. That is, we are not interested in other types of absenteeism such as chronic illness, broken bones, or other types of absence which could not even logically be related to drinking water.

Obviously; any discussion concerning increases in operator salaries implies that it is the lower salaries which are most in need of improvement.

The second assumption is that across a large population, the occurrence of acute illness in adults is a responsibly stable proportion of its occurrence in children. This does not imply a one-to-one relationship nor does it imply that the constant of proportionality is the same for any specific adult-child pair.

Finally, we assume that this constant is measured by the ratio $\dfrac{A_i}{C_i}$ where A_i is the average, across the population being studied, of annual adult days lost from work with infective and parasitic diseases and digestive system conditions. C_i is the counterpart for children's days lost from school. We have eliminated respiratory conditions from the ratio, not because it is impossible to communicate them via drinking water, but because the others are those most commonly associated with drinking water.

The National Health Survey of 1969 found that there were 100.9 days of work lost per 100 employed due to the above described acute conditions. They found that there were 163.8 days lost from school per 100 students due to the above conditions. A check across an SMSA vs non-SMSA population showed that despite different rates of bed disability with acute illnesses, A_i/C_i ratios differ by a maximum of .006. Checks across different age groups show that in the 15-14 vs 45-64 age brackets, the ratios are identical. Given the other error in the data, we believe this difference to be insignificant. Given that, we will use the ratio $\dfrac{100.9}{163.8}$ = .616 as the constant of proportionality. This means that we expect that the percentage of estimated water-related absenteeism in the adult population is .616 times that of children in elementary school.

Additionally, the survey found that, excluding hospital care, dental bills, and special expenses, persons with no chronic illnesses spent an average of $69 per year on doctors and medicine. The non-respiratory and non-injury acute illnesses account for 30 percent of the total disability

days. [58] Thus, we estimate that the combined doctor and medicine expenses per person per year for the acute illnesses which we are considering is roughly $20.7.

The labor force in Vermont to which this study is applicable numbers 153,094 with a total income of 873 million dollars per year. Each one point increase in average operator salary index should reduce absenteeism by .34 percent of the total worked days. Annually for the total state, this would mean a savings of $2,980,000 in wages or a savings of roughly $20 per working person per year for each average increase of one point on the operator salary index. Using the rough conversion factor of $2500 per index point, the savings are $1,190,000 or $8.00 per worker for each $1000 average increase in operator salary.

In the area of interconnections with unprotected systems, the savings is considerably less. For each ten percent average improvement in the interconnections, we can expect a reduction in absenteeism of .02 percent. Statewide this amounts to a savings of $170,000 per year in wages or slightly more than $1 per worker.

Total medical costs in the illness with which we are concerned are $9.6 million dollars.

Since each $1000 average salary increase reduces absenteeism by .136 per cent of the total worker days, this accounts for $\frac{.00136}{.25} = .055$ of the illnesses in the acute non-respiratory, non-injury category. This amounts to a savings in medical costs of $545,000 per year per $1000 average salary increase. Similar calculations show a savings of $77,000 per year per ten percent improvement in interconnections with unprotected systems.

The total savings (salaries plus medical costs) are 1.7 million dollars per year per $1000 average increase in operator salaries and $247,000 per year per ten percent improvement in the interconnections.

New York

The New York study examined non-farm communities and suburbs in Nassau, Rockland, Suffolk and Westchester counties. We also examined New York City, but for reasons explained above, had to leave it out of the final analyses. The applicable population in the above-mentioned com-

---

58)National Center for Health Statistics, "Personal Health Expenses - Per Capita Annual Expenses, United States: July-December 1972," U. S. Department of Health, Education and Welfare, Public Health Service, Series 10, Number 27, February 1966.

munities number 3.67 million persons and 645,000 elementary school students. The final regression employed an index of environmental quality formed by dividing the index of operator salary by the annual average particulates in the air. The range of the index is from .12 to .01. The regression coefficient for the index is 11.4 with a standard deviation of 4.4. Thus, for each .01 change in the index, absenteeism will drop by .11 percent of the total population.

With the 118 million student days under consideration, this means that for each improvement of .01 in the environmental index, absenteeism in these four counties will drop by 133,000. The one and two standard deviation unit ranges respectively, are 81,000 to 185,000 and 31,000 to 235,000 student days per .01 improvement. Interpreting the salary index of $2500 per increment, an average improvement of $1000 in the salaries of the operators will, [59] holding air pollution constant, reduce absenteeism by 71,500.

The New York population which is affected by the study, totals 3.7 million persons with a working force of 1.5 million and a total income of $15.6 billion. Using the environmental index comprised of the index of operator salary divided by average particulates, the adjusted regression coefficient for the working population is 7.25. Thus, for every .01 increase in the index, worker absenteeism will fall by .07 percent of the total population. For the population under consideration this means a savings in lost wages of $10,900,000 per average index improvement of one percent. This averages $7.5 per worker in the four county areas surrounding New York City.

Assuming average and unvarying particulate counts, this translates to a total savings of $5.95 million dollars per year per $1000 average increase in operator salary.

Medical expenses for the illnesses under consideration in the four county area total 94 million dollars per year. Each $1000 average increase in operator salaries under the above air pollution conditions will produce a $\frac{.038}{2.5}$ percent reduction in absenteeism due to our subject of acute diseases. This amounts to a savings of 1.42 million dollars per year. Thus, the total savings per $1000 increase in operator salaries is 7.37 million dollars per year.

---

59) This is the empirically observed average.

Six Area Group-Rural

In the other six metropolitan areas examined in the study, we divided our sample into urban and rural sectors. The applicable rural population affected is 480,000 with 82,000 elementary students. Both frequency of testing of finished water and operator salaries are strong predictors.

The regression coefficient of testing of finished water is 1.04. The variable is scaled into intervals roughly equal to one year. The regression coefficient for operator salary is 1.0.

There are 14.8 million student days covered by the sample. For each average increase in the frequency of testing finished water we can expect an increase of 160,000 student days attendance in these areas. The one standard deviation range would be 113,000 to 208,000 and the two standard deviation range would be 65,000 to 255,000.

The expected reduction in absenteeism resulting from increased operator salary is 148,000 student days per one point average increase in the salary index. This translates roughly to an increase of 63,000 student days attendance per $1000/year increase in the average salary. The one and two standard deviation ranges are 31,000 to 95,000 and 0 to 127,000, respectively.

The population of this six area group in the study is 479,000 with a labor force of 170,000 and a total income of $1.2 billion dollars. The primary predictor variable is the frequency of testing finished water with an adjusted regression coefficient for the adult population of .64. Thus, for each one year average reduction in the frequency of testing,absenteeism will be reduced by .64 percent of the population. This amounts to $7.2 million or approximately $45 per worker.

The second significant variable is the salary of the operator which has an adjusted regression coefficient of .616. This would lead us to predict a savings of 2.9 million dollars annually for each $1000 average increase in operator salaries. This is a savings of roughly $18 per worker per year.

The related medical expenses for this group is 9.9 million dollars per year.

The medical savings per average one year improvement in testing frequency is 2.6 million dollars per year. A $1020 average salary increase will produce a savings of $970,000.

The total savings for improving the testing procedures are 9.8 million dollars per one year increased frequency. The total savings resulting from a $1000 operator salary increase is 3.9 million dollars.

National Extrapolation

The above calculations have been based on a limited sample of 8.4 million persons, concentrated within eight relatively small geographical areas. The original CWSS study of nine areas made no claims that the sample was representative of a larger universe. The sample was reasonably diverse geographically although only three of the nine were west of the Mississippi River.

It included a diversity of economic and cultural characteristics and covered a wide spectrum of community sizes. Despite this, they did not choose a scientific sample from which we can generalize to the total population with any known degree of statistical confidence. Thus, it is only prudent for us to advise the reader that the following national estimates are based only on the unsupported assumption that the sample has drinking water vs. health characteristics which are similar to those of the entire nation.

We will compute the rational extrapolations from those of our models: (1) the general model, (2) the Vermont model, and (3) the six area rural model. The six area-urban/suburban model is too tenuous to permit such extrapolations. National extrapolations from the New York model are seriously hampered by the strong interactions with air pollution.

The Vermont findings of the effect of interconnections with unprotected systems is difficult to generalize given that the CWSS found that across these total samples only six percent of the cases have this problem. We will, therefore, not extrapolate this portion of the Vermont results.

The national estimates are straight line extrapolations using the total number of enrolled elementary students as the basis for elementary school absenteeism, total wages paid as the basis for total lost wages saved and total population for special medical costs.

Extrapolation from the General Model

Recalling that in the general model, we found that the idea of operator salary times the frequency of testing was the strongest water related variable. The regression coefficient was .08, meaning that for every 1.0 reduction in the index value,absenteeism would be reduced by .08 percent ± .08 at the .05 confidence level. With 5.4 billion student days nationally,

an improvement of 1.0 in the index would reduce elementary school absences by 4.3 million student days.

Expanding this to the adult population in 1969 there were 57.9 million persons in the private sector of the labor force excluding rural farm workers. These persons were employed for approximately 15 billion days and earned nearly 390 billion dollars. Each 1.0 reduction in the index of salary times frequency of testing would reduce absenteeism due to illness by 7.4 million days ± 7.1 million days resulting in a savings of 194 million dollars ± 186 million dollars from wages alone. The total of 7.4 million days of worker illness plus 4.3 million days of school illnesses means a reduction of 11.7 million days in sub-clinical illness apparently resulting from drinking water.

The National Health Survey reports indicate that illnesses in this general category cost an average of $5.46 per illness, including medicine and physicians. We do not know the average length of illness in this category. Assuming, however, that it is close to 1.0 days per illness, and accepting the very gross assumption of the national extrapolation, let us estimate a cost of $5.00 medical costs per day of sub-clinical illness. Given one estimate of 11.7 ± 11.2 million days illness, the medical savings are $58.5 million dollars ± 56 million dollars. From the general model, therefore, we would estimate savings of 252 million dollars per year ± 242 million dollars for each 1.0 improvement in the index.

It should be remembered that the composition of the index makes substantive interpretation somewhat difficult. Since better (more highly paid) operators could be expected to engage in better testing procedures, we cannot expect that realistically we can hold testing constant while improving operator quality (salary). Similarly we probably do not expect a general improvement in practice until municipalities decide to improve what are clearly inadequate salaries. Since the relationship in the general model is greatly strengthened by the multiplicative term, the reader must realize that a 1.0 increase implies a probable improvement in both elements.

Vermont Extrapolation

Although we did not exclude urban areas from the Vermont studies, those which were included were so small relative to the other large metropolitan areas in the country, that it appeared more reasonable to extrapolate only to the rural non-farm portions of the United States. We are extrapolating to a working population of 1.2.4 million persons earning 70.9 billion dollars.

The Vermont analyses applied to 409,000 persons and concluded that an average increase of $1000 per year in operator's salaries would reduce elementary school absenteeism by 27,000 days per year and save 1.7 million dollars per year in wages and medical expenses.

To the extent that these findings are applicable to all rural non-farm communities there would be a nationwide reduction of 3.1 million elementary school absentee days per year per $1000 average increase in the salary of the operators. This is based on a nationwide rural non-farm elementary school population of 7.7 million students.

The reduction in lost wages would be 138 million dollars per year and the savings in medical expenses would be 50 million dollars. Thus, using the Vermont model the total savings to the rural non-farm population per average $1000 increase in operator salary would amount to 188 million dollars per year.

Six Area Grouping   Rural Extrapolation

In the six area rural grouping we found that operator salary was a significant variable with a regression coefficient of nearly twice that of Vermont.

Applying those findings to the total U. S. rural non-farm population, we would predict a total reduction of 5.5 million absentee days from elementary schools per $1000 average increase in operator salary. We would also predict a savings of 246 million dollars per year in lost wages and another 80 million dollars per year in medical expenses.

In the six area groupings, we found that the frequency of testing finished water was also quite significant. In the directly affected 480,000 population, we predicted a reduction of 160,000 days elementary school absenteeism and a savings of 7.2 million dollars in wages and another 2.6 million in medical expenses for each 1 year average increase in the freqency of testing of finished water

Extending these findings nationwide to all rural non-farm communities, we would expect a reduction of 15 million days elementary school absenteeism, savings of 599 million dollars in lost wages and 230 million dollars in medical expenses.

If we were to assume a testing improvement of frequency/month for each, rather than frequency/year we would predict a reduction of 1.3 million days elementary school absenteeism, savings of 49 million dollars in lost wages and 19 million dollars in medical expenses.

Recalling the fact that improvement in salary and improvement in testing are likely to partially co-occur, it is not reasonable to compute a joint savings of salary improvement and testing frequency.

Again we would urge the reader to be cautious in interpreting these national extrapolations. The extrapolations are best estimates given existing information but we have only limited basis for making any statements of statistical confidence.

Conclusions

In concluding this report, there are several points we would like to make. The first is that the existence of a relationship between drinking water and sub-clinical absenteeism in elementary schools is strongly supported by the study. The results are highly statistically significant consistent with our hypotheses, and stable across different sub-samples. Even in the weakest cases, partial correlations were above .33, thus accounting for more than ten percent of the unexplained variance.

The problem of operator salary is one which can easily be misconstrued. Therefore, it is worth reemphasizing a point made earlier. Although operator's salaries are strong predictors of absenteeism, we would caution against the conclusion that pay raises will immediately improve community health. The astoundingly low salaries have resulted in the recruitment of generally poorly educated, undertrained operators. While in time, higher salaries will make possible the selection of higher caliber operators, in the short term, the salaries would have to be supplemented by vigorous operator improvement programs if any significant result is to be expected.

Therefore, we computed benefits on the basis of increments in improvements in the various measures of water quality. The various increments predict reductions of 3% to 10% of the digestive and non-respiratory illnesses. In the specific areas studied, this translated to reductions in elementary school absences on the order of 300,000 to 600,000 absences per year per incremental improvements. Converting this to worker absenteeism using a constant of proportionality of .62, this becomes a savings in the order of 23 million dollars per year per increment.[60] For the nationwide sample, we would cautiously estimate that elementary school absences would be reduced on the order of magnitude of 11-12 million absences per year per incremental improvement. The wages and medical expenses saved were in the order of 250 million dollars per year per incremental improvement.

---

60) These ranges are very broad and should be interpreted only as "ball park" estimates. The more precise estimates which are presented in the body of the report, require breakdown by categories of drinking water problem and geographic area.

In performing the study, we made every attempt to be conservative in our estimates. The medical cost figures, for example, assume that there are no respiratory diseases resulting from drinking water. If they were transmitted in a manner similar to gastro-intestinal illness, the medical cost figures would be more than twice what we have estimated. At all other points in the research program, when we were faced with two alternatives, one of which would tend to inflate the results and the other of which would tend to deflate them, we systematically chose the more conservative.

We believe that the results indicate that if improvements are made in the quality of drinking water, substantial savings will be realized. As the CWSS report suggested and our study has confirmed, the problems are most severe in the rural non-farm communities.[61] There the economic savings alone would be most dramatic both because the effects are relatively large and the costs of resolving the problem are relatively small. In most cases, upgrading the personnel and procedures appears to be all that is needed.

---

61) Rural farm populations are excluded from the conclusions because they were not included in either study.

SECTION VI

BIBLIOGRAPHY

Alter, Milton. "The Distribution of Multiple Sclerosis and Environmental Sanitation", Israel Journal of Medical Sciences, 1972, 99-105.

Barrett, Larry B. and Waddell, Thomas E. "Cost of Air Pollution Damage: A Status Report," Environmental Protection Agency, Research Triangle Park, North Carolina, February 1973.

Berg. G. (ed.). Transmission of Viruses by the Water Route, New York: Inter science Publishers (1967).

Borneff, J. Public Health Aspects of Viruses in Water, N. A. Clarke (ed.) Cincinnati, Ohio, 1 (1970), 1.

Chapman, Robert S. et al. "Chronic Respiratory Disease Symptom Frequency in Late Adolescence, In Parents of School Children and in Elderly Persons," Presented at the American Medical Association's Air Pollution Medical Research Conference, Chicago, Illinois, October 1972.

Chelsky, Morris. "The Propagating Rate: A Method for Interpreting Infectious Disease Incidence". American Journal of Public Health, September 1969, 1661.

Chin, T. D. Y., W. H. Mosley, S. Robinson and C. R. Gravelle, "Detection of Enteric Viruses in Sewage and Water, Relative Sensitivity of the Method", Transmission of Viruses by the Water Route, G. Berg (ed.), New York: Interscience Publishers (1967).

Clarke, N. A. "Viruses in Water," Journal of the American Water Works Association, 61:10 (1969), 491-494.

Cohen, A., et al. "Asthma and Air Pollution from a Coal-Fueled Power Plant," Asthma and Air Pollution, American Journal of Public Health, September 1972, 118 1-1.188.

Craun, Gunther and Leland McCabe. "Waterborne Disease Outbreaks, 1961-1970", Presented at the Annual Meeting of the American Waterworks Association, Colorado, June 1971.

Dauer, C. C. "1951 Summary, Food and Waterborne-Disease Outbreaks". Public Health Reports, 1952, 1089.

Dohan, F. C., G. S. Everts, and R. Smith. "Variations in Air Pollution and the Incidence of Respiratory Disease". Journal of the Air Pollution Control Association, 1962, 418-436.

Eliassen, R. and Cummings, R. H. "Analysis of Waterborne Outbreaks, 1938-45", Journal of the American Waterworks Association, 1948, 509.

Enviro Control, Inc. "National Assessment of Trends in Water Quality" Final Report submitted to Council on Environmental Quality, June 1972.

Environmental Protection Agency, "National Water Quality Inventory Report", Office of Water Planning and Standards, November 1973.

Fairchild, Glen A., Judy Roan, and James McCarroll. "Atmospheric Pollutants and the Pathogenesis of Viral Respiratory Infection", Arch. Environmental Health, September 1972, 174-182.

French, J. G., et al. "The Effect of Sulfur Dioxide and Suspended Sulfates on Acute Respiratory Disease in Children and Adults in Four Areas of the United States," Presented at the AMA's Air Pollution Medical Research Conference, Chicago, October 1972.

Glass, A. C. and K. H. Jenkins. "Statistical Summary of Municipal Water Facilities in the United States", USPHS Pub. No. 1039, January 1958.

Gorman, A. E., and Wolman, A. "Waterborne Outbreaks in the United States and Canada, and Their Significance". Journal of the American Waterworks Association, 31:225, 1939.

Harris, Robert H. and Edward M. Brecher, and the Editors of Consumer Reports. "Is the Water Safe to Drink?" Consumer Reports, June 1974, 436-443.

Harvard University, "The Economics of Water Supply and Quality". Project No. 16110 DTF. Environmental Protection Agency, February 1971.

Hill, William F., Elmer W. Akin and William H. Benton. "Detection of Viruses in Water: A Review of Methods and Application", Proceedings of the 13th Water Quality Conference, Illinois 1971.

Kramer, H. P. "Research Need for Drinking Water", Journal of the American Waterworks Association, 57:1 (1965). 3-9.

Krishnaswami, S. K. "Health Aspects of Water Quality", American Journal of Public Health, Vol. 61:11, November 1971, 2259-2268.

Lackner, Jack. "Safe Drinking Water Act of 1973: Estimated Benefits and Costs", Environmental Protection Agency, Office of Planning and Evaluation, January 1973.

Lave, L. B. and E. P. Seskin. "Air Pollution and Human Health", Science 169 (3947): 723-733, August 21, 1970.

_____. "An Analysis of the Association between U.S. Mortality and Air Pollution", Carnegie-Mellon University, September 1970.

Liu, Oscar C, M. D. Private Communication to Dennis Tihansky, Environmental Protection Agency, August 29, 1972.

Lyon, W. A. "Water and Health - Are We Concerned Enough?" Journal of Sanitary Engineering Division, October 1970.

McDermott, J. H. "Virus and Water Quality: Occurrence and Control - Conference Summary", Environmental Protection Agency, 1971.

McKee, J. E. and H. W. Wolf. "Water Quality Criteria". California State Water Quality Control Board, Pub. No. 3-A, 1963.

McLean, D. M. et al "Virus Dispersal by Water" Journal of the American Waterworks Association, 1966, 920.

Melnick, J. L. "Detection of Virus Spread by the Water Route", Proceedings of the 13th Water Quality Conference, Illinois, 1971.

Metcalf, T. G. "Biologic Parameters in Water Transmission of Viruses" Proceedings of the 13th Water Quality Conference, Illionis, 1971.

Mosley, James W. "Transmission of Viral Diseases by the Water Route." New York: J. Wiley and Sons, 1965.

New York, State of. "Medical Fee Schedule", Workmen's Compensation Board, July 1960.

Neri, L. C., J. S. Mandel, D. Hewitt. "Relation Between Mortality and Water Hardness in Canada", The Lancet, April 29, 197'2, 931-934.

Ongerth, Henry, et al. "Public Health Aspects of Organics in Water" Journal of the American Water Works Association, July 1973, 495-498.

Petersen, Norman J. and Virginia D. Hines. "The Relation of Summer-time Gastroentestinal Illness to the Sanitary Quality of the Water Supplies in Six Rocky Mountain Communities", American Journal of Hygiene. 1960,

Poynter, S. F. B. "The Problem of Viruses in Water", Water Treatment and Examination 1968, 187-204.

Purdue University, "The Utilization of Health Services: Indices and Correlates", Health Services Research and Training Program, Indiana, 1972.

Ridker , Ronald G. Economic Costs of Air Pollution, New York: Frederick A. Praeger, 1967.

Schafer, E. Public Health Aspects of Viruses in Water, N. A. Clarke, (ed.), Cincinnati, Ohio, 1 (1970), 14.

Schroeder, H. A. "Municipal Drinking Water and Cardiovascular Death Rates", Journal of the American Dietetic Association 195: 2, 1966, 125-129.

Shakman, Robert A. "Nutritional Influences on the Toxicity of Environmental Pollutants," Arch Environmental Health, February 1974, 105-113.

Shy, Carl M. and John F. Finklea. "Air Pollution Affects Community Health", Environmental Science and Technology, Volume 7: 3, March 1973, 204-208.

Shy, C. M., et al. "An Overview of CHESS". In Health Consequences of Sulfur Oxides : A Report from CHESS, 1972.

Smith, L. S. et al. "Public Health Aspects of Water Pollution Control", Water Pollution Control, 1969, 544-549.

Social Security Administration, "Medical Care Costs and Prices: Background Book", Department of Health, Education and Welfare.

Stevenson, A. H.   "Studies of Bathing Water Quality and Health". American Journal of Public Health, 43:5, 1953, 529-538.

Thompson, Donovan, J. et al.   "Health and the Urban Environment: Air Pollution, Weather and the Common Cold, "American Journal of Public Health 60:4, p. 731-739.

Tihansky , D.   "Economic Damages from Residential Use of Mineralized Water Supply", Water Resources Research, April 1974.

USPHS.   Drinking Water Standards - 1962. P. H. S. Publication No. 956, 1963, 61.

Viraraghavan, T.   "Water Quality and Human Health".   Journal of the American Water Works Association, October 1973, 647-650.

Weibel, S. R. et al.   "Waterborne-Disease Outbreaks, 1946-60", Journal of the American Water Works Association, August 1964, 947-958.

William, D. B.   "Viruses in Water Supply - Are Present Treatment Practices Adequate?" Paper Presented at the Ontario Sect. Annual Conference. Toronto, 1971.

APPENDIX A

CALIFORNIA RESULTS

N = 32   Significance of Regression   =.99

Absenteeism = 6.47 - .09 Sw + 2.6E - 1.45F

                                                      Significance

Sw =   Number of water samples
              collected/month   .95
E =   Education of operators   .90
F =   Frequency of testing of water   .95

The dependent variable in the California study supposedly was unexcused absences from elementary schools. Because of this we had originally thought that it would be a good test of the null hypothesis (i.e. no relationship between water quality and training), Unfortunately, we were reminded, lower income children and children of transit parents (numerous in the San Bernardino-Riverside area) frequently do not have excuses even when legitimately ill. Therefore, the measure is quite "dirty." By this we mean that we do not know with certainly what we are measuring.

The results of the regression reflect this confusion. The two operations variables measuring frequency of sample collection and frequency of testing are statistically significantly linked in the hypothesized direction to absenteeism as though it were truly measuring illness. The education of the operators is strongly linked in the <u>opposite</u> direction. We would conclude from the latter that for each incremental improvement in education of the operators, truancy (or illness) would increase by 2.6 percent of the population. These findings make sense only if we interpret the dependent variable as truancy. In that event, truancy, which tends to be higher in urban areas than in rural areas, is positively correlated with operator education which also tends to be higher in urban areas than in rural areas. If we assume that the dependent measure captures variance both of truancy and of illness, then the results are somewhat reasonable,   In that case, the two operations variables are capturing the illness portion of the variance while education of the operators is capturing the truancy portion.

71

Admittedly, the argument is less than convincing, and, to some extent, that supports the original purpose of the test of the California data. That is, because of the dependent variable, it is the only area in which we have to make such strained arguments defending the linkage between water quality and absenteeism. In all other areas the drinking water quality results were much more sensible and straight-forward. We think that this demonstrates that the methodology itself does not necessarily produce positive results which are statistically significant and consistent with one hypotheses.

The Profile of Medical Practice

Average Fee for Initial Office Visits
by Census Division and Specialty, 1970

Census Division	General Practice
Total	$ 8.46*
New England	9.19
Middle Atlantic	8.66
East North Central	8.18
West North Central	7.27
South Atlantic	8. 56
East South Central	7.13
West South Central	7.21
Mountain	7.78
Pacific	10.67
Average	7.49

*Based on 852 observations

Notes

The less populated regions (with relatively low concentrations of physicians) experienced lower levels of fees for certain office visits than the more populated regions (with relatively high concentrations of physicians).

Variations in fees presumably reflect differences in the orientation and training of physicians in various specialties, as well as patient demand for their services and other factors.

In this study fees have been compared for a very few general categories of service. Medical specialists differ widely in the types of service they provide, and the demand for these services differ among geographic regions. Specific comparison of fees are therefore difficult. At the very least, explanations for the marked variations require a great deal more knowledge than is contained in the available data. Given these considerations, assessment of variations and general trends in physicians' fees should be approached with caution.

Source: Sixth Periodic Survey of Physicians. Nov. 1970. Center for
Health Services Research to Development, AMA.